

Copyright

by

Craig R. Bush

2007

**COMMITTEE CERTIFICATION OF APPROVED VERSION**

**The committee for Craig Randall Bush certifies that this is the approved version of the following dissertation:**

**FUNCTIONAL GENOMIC ANALYSIS OF PPAR-GAMMA IN HUMAN COLORECTAL CANCER CELLS**

**Committee:**

---

**E. Aubrey Thompson, Ph.D, Supervisor**

---

**Bruce A. Luxon, Ph.D, Supervisor**

---

**E. Brad Thompson, MD**

---

**Allan R. Brasier, MD**

---

**Larry A. Denner, Ph.D**

---

**Rudy Guerra, Ph.D**

---

Dean, Graduate School

**FUNCTIONAL GENOMIC ANALYSIS OF PPAR-GAMMA IN  
HUMAN COLORECTAL CANCER CELLS**

**By**

**Craig Randall Bush, B.S.**

**Dissertation**

**Presented to the Faculty of The University of Texas Graduate School of  
Biomedical Sciences at Galveston  
in Partial Fulfillment of the Requirements  
for the Degree of**

**Doctor of Philosophy**

**Approved by the Supervisory Committee**

**E. Brad Thompson, MD  
E. Aubrey Thompson, Ph.D  
Bruce A. Luxon, Ph.D  
Allan R. Brasier, MD  
Larry A. Denner, Ph.D  
Rudy Guerra, Ph.D**

**December, 2006  
Galveston, Texas**

**Key words: systems biology machine-learning**

**© 2007, Craig R. Bush**

**DEDICATION**

**To my wife, Lu**

## ACKNOWLEDGEMENTS

I wish to first and foremost thank both of my mentors, Drs. E. Aubrey Thompson and Bruce A. Luxon for their guidance and wisdom through this long and difficult exercise. Aubrey, here's to sitting out on the lanai, looking out over that salt marsh, and pondering the wonders of the world. Bruce, thank you for taking a chance on me and then taking another. I will never forget that.

I would like to acknowledge the other members of my supervisory committee, Drs. E. Brad Thompson, Allan R. Brasier, Larry A. Denner, and Rudy Guerra for their advice and insight.

Special thanks to members of my laboratories for all of their hard work. At the Mayo Clinic in Jacksonville, Florida: Drs. Brian Necela and Weidong Su, Mrs. Nancy Rutkoski and Miss Jennifer Havens. At the University of Texas Medical Branch in Galveston, Texas: Mrs. Mala Sinha. Likewise, I wish to also thank Drs. Panagiotis Anastasidasis and Masahiro Yanagisawa and Ms. Michelle Guigneaux for their technical assistance.

I would also like to thank my wife Lu for her patience and encouragement. I simply do not think I could have completed this work without her.

I wish to extend my gratitude to the Keck Center for Computational Biology at Rice University for granting me a three year pre-doctoral National Library of Medicine fellowship and Sankyo Ltd. (Tokyo, Japan) for financial support and use of their compound.

Finally, I would like to thank Dr. David Niesel for taking time out of his busy schedule to talk to me that one afternoon in the Summer of 2000.

# FUNCTIONAL GENOMIC ANALYSIS OF PPAR-GAMMA IN HUMAN COLORECTAL CANCER CELLS

Publication No. \_\_\_\_\_

**Craig Randall Bush, Ph.D**

**The University of Texas Graduate School of Biomedical Sciences at Galveston, 2007**

**Supervisors: E. Aubrey Thompson and Bruce A. Luxon**

The gamma isoform of peroxisome-proliferator activated receptor (PPAR $\gamma$ ) is a member of the super family of nuclear hormone receptors and shows much promise as a chemopreventative and therapeutic target for colorectal cancer. Activation of PPAR $\gamma$  by thiazolidinediones (TZDs) inhibits proliferation and induces differentiation in human colon cancer cells. RS5444, a novel TZD, is a high affinity and high specificity ligand for PPAR $\gamma$ . We have shown that RS5444 markedly reduced the proliferation of MOSER S human colorectal cancer cells under anchorage dependent and independent conditions. The inhibitory effect of RS5444 was irreversible. RS5444 also significantly repressed the invasive phenotype, but not motility, of these tumor cells.

Towards elucidating the activated PPAR $\gamma$  controlled genomic program responsible for these observed phenotypes, functional genomic analysis was performed using a two-class longitudinal microarray data set in the presence and absence of RS5444. Differential expression of genes was calculated using an empirical Bayesian modification to the multivariate HotellingT2 score. We have demonstrated this statistical machine learning technique to be superior in controlling type II error in our dataset than more commonly used algorithms for two-class analysis. Likewise, through the use of

several bioinformatics techniques, including frequency-based pathway and ontology analysis, we found a yet unappreciated tumor-suppressing network involving a feedback mechanism between PPAR $\gamma$ , DSCR1 and calcineurin-mediated signaling of NFATc in colorectal cancer cells. To this end, we have demonstrated a direct connection between NFATc and DSCR1 in MOSER S colorectal cancer cells. Likewise, we have demonstrated a correlation between the sensitivity of PPAR $\gamma$  in other colorectal cancer cells, and the messenger abundance of DSCR1. Finally, we have demonstrated that knockdown of DSCR1 messenger obviates the phenotypic effects of activated PPAR $\gamma$  *in vitro*.

To our knowledge these data represent, for the first time, a network between PPAR $\gamma$ , DSCR1, and NFATc signaling in the context of tumor-suppressor activity. This preliminary evidence is consistent with our working hypothesis that an oncology patient's receptiveness to TZD treatment may be largely dependent on the specific tumor's endogenous abundance of DSCR1. We believe without a critical endogenous level of DSCR1, activated PPAR $\gamma$  may revert to a tumor-activator instead of a tumor -suppressor.

# TABLE OF CONTENTS

<b>LIST OF TABLES</b>	<b>X</b>
<b>LIST OF FIGURES</b>	<b>XI</b>
<b>LIST OF ABBREVIATIONS</b>	<b>XIV</b>
<b>GLOSSARY</b>	<b>XVII</b>
<b>CHAPTER 1: BACKGROUND</b>	<b>1</b>
<b>SYSTEMS BIOLOGY</b>	<b>1</b>
The Systems Biology Paradigm Shift .....	1
Functional Genomics' Role in Discovery Research .....	11
Systems biology and reverse-engineering complexity .....	13
Examples of tools that have been used to solve systems problems .....	14
<b>PEROXISOME PROLIFERATOR-ACTIVATED RECEPTORS</b>	<b>24</b>
Nuclear hormone receptors .....	24
PPAR family .....	25
PPAR $\gamma$ .....	28
PPAR $\gamma$ and colorectal cancer .....	33
<b>DSCR1, CALCINEURIN AND NFAT</b>	<b>36</b>
DSCR1 as a tumor suppressor .....	36
NFATc as a tumor promoter .....	37
Calcineurin .....	38
The signaling axis of DSCR1-Calcineurin-NFATc .....	39
<b>CHAPTER 2: MATERIALS AND METHODS</b>	<b>44</b>
<b>CHEMICAL REAGENTS</b>	<b>44</b>

<b>PLASMIDS AND VECTORS</b>	<b>44</b>
<b>CELL LINES</b>	<b>45</b>
<b>WESTERN BLOTTING</b>	<b>47</b>
<b>TRANSIENT TRANSFECTION ASSAY</b>	<b>48</b>
<b>ANCHORAGE-DEPENDENT AND –INDEPENDENT CELL PROLIFERATION</b>	<b>49</b>
<b>COLONY FORMATION ASSAY</b>	<b>50</b>
<b>MIGRATION AND INVASION ASSAYS</b>	<b>51</b>
<b>SHRNA GENE SILENCING</b>	<b>52</b>
<b>RNA ISOLATION</b>	<b>52</b>
<b>QUANTITATIVE REAL-TIME PCR (QPCR)</b>	<b>54</b>
<b>ESTIMATION OF RS5444’s EC50</b>	<b>56</b>
<b>MICROARRAY ANALYSIS</b>	<b>57</b>
<b>QUALITY CONTROL AND PREPROCESSING OF AFFYMETRIX MICROARRAY PROBE-LEVEL DATA</b>	<b>58</b>
<b>STATISTICAL MODELS FOR DIFFERENTIAL EXPRESSION OF LONGITUDINAL MICROARRAY DATA</b>	<b>60</b>
Non-parametric model .....	60
Polynomial Regression Model .....	62
Empirical Bayesian Model.....	64
<b>GENE ONTOLOGY ANALYSIS</b>	<b>68</b>
<b>CHAPTER 3: RESULTS</b>	<b>70</b>
<b>BIOACTIVITY OF RS5444, A HIGH AFFINITY PPAR<math>\gamma</math> SPECIFIC AGONIST</b>	<b>70</b>
<b>CHARACTERIZATION OF MOSER S HUMAN COLORECTAL CANCER CELLS</b>	<b>76</b>
<b>CHARACTERIZATION OF MOSER S CELLULAR PHYSIOLOGY UPON PPAR<math>\gamma</math> ACTIVATION BY RS5444</b>	<b>77</b>

RS5444 did not alter PPAR $\alpha$ receptor abundance or transcriptional activity .....	77
Activation of PPAR $\alpha$ caused inhibition of MOSER S anchorage-dependent and – independent growth.....	80
The anti-proliferative effect of RS5444 was irreversible .....	87
Activation of PPAR $\alpha$ by RS5444 inhibited invasiveness of MOSER S cells.....	87
<b>FUNCTIONAL GENOMICS ANALYSIS OF RS5444 EFFECT ON MOSER S COLORECTAL CANCER CELLS</b>	<b>92</b>
<b>EXPERIMENTAL DESIGN OF MICROARRAY TIME COURSE EXPERIMENT</b>	<b>93</b>
<b>QUALITY CONTROL AND PREPROCESSING OF AFFY MICROARRAY DATA</b>	<b>96</b>
Quality Control Step: Visual Inspection of CEL images.....	96
Determination of Preprocessing Algorithm .....	99
Quality Control Step: Unsupervised Clustering .....	106
Probe Set Filters .....	106
Analysis of Performance of Differential Expression Models .....	109
Differential Expression Analysis by empirical Bayes HotellingT2 score .....	113
Descriptive statistics of GCRMA summarized expression.....	115
Time course kinetic classes.....	124
Ontology analysis of gene expression.....	124
IPA infers a DSCR1-calcineurin-NFATc signaling axis .....	143
<b>CHAPTER 4: DISCUSSION</b>	<b>173</b>
<b>APPENDIX A</b>	<b>189</b>
<b>APPENDIX B</b>	<b>252</b>
<b>REFERENCES</b>	<b>256</b>
<b>VITA</b>	<b>291</b>

## LIST OF TABLES

	<b>Page</b>
<b>Table 1:</b> Primers for MOSER S cells	<b>46</b>
<b>Table 2:</b> DSCR1 shRNA MISSION lentiviral clone candidates	<b>53</b>
<b>Table 3:</b> qPCR Applied Biosystems TaqMan® Assays	<b>55</b>
<b>Table 4:</b> Observed fold change of PPAR $\gamma$ marker PDK4 by TZD dose response curve	<b>74</b>
<b>Table 5:</b> Fitting parameters for single binding affinity model assuming saturation	<b>75</b>
<b>Table 6:</b> Lowess scores of preprocessing methods	<b>103</b>
<b>Table 7:</b> Statistical power of differential expression models by boot-strapping with replacement	<b>112</b>
<b>Table 8:</b> The top twenty positively and negatively regulated probe sets	<b>119</b>

## LIST OF FIGURES

	<b>Page</b>
<b>Figure 1:</b> Illustration of autonoma in Descartes' thesis Descourses, 1637	<b>4</b>
<b>Figure 2:</b> Serial versus parallel signal transduction	<b>8</b>
<b>Figure 3:</b> Common structural features of nuclear hormone receptors	<b>27</b>
<b>Figure 4:</b> Alternate isoforms of the human PPAR $\gamma$ gene	<b>30</b>
<b>Figure 5:</b> Calcium-Calcineurin-NFATc Signaling	<b>41</b>
<b>Figure 6:</b> Bioactivity of a high affinity PPAR $\gamma$ agonist RS5444	<b>73</b>
<b>Figure 7:</b> Characterization of the APC gene and its expression in MOSER S human colon cancer cells	<b>79</b>
<b>Figure 8:</b> RS5444 did not alter PPAR $\gamma$ expression level or transcriptional activity in MOSER S cells	<b>82</b>
<b>Figure 9:</b> RS5444 inhibited MOSER S cell proliferation under anchorage-dependent conditions	<b>84</b>
<b>Figure 10:</b> RS5444 inhibited MOSER S cell proliferation in soft agar	<b>86</b>
<b>Figure 11:</b> RS5444's anti-proliferative effect was irreversible	<b>88</b>
<b>Figure 12:</b> Activation of PPAR $\gamma$ by RS5444 inhibited invasion, but not migration, of MOSER S cells	<b>91</b>
<b>Figure 13:</b> Experimental design of microarray time course	<b>95</b>
<b>Figure 14:</b> Inspection of microarray images for artifacts	<b>98</b>
<b>Figure 15:</b> Bland-Altman plot of preprocessing methods	<b>101</b>
<b>Figure 16:</b> Averages of lowess regression scores at each time point	<b>105</b>
<b>Figure 17:</b> Principal Components Analysis (PCA) of microarray time course	<b>108</b>

<b>Figure 18:</b>	The top and bottom ranked differentially expressed probe sets within the MB-statistic cutoff	<b>117</b>
<b>Figure 19:</b>	A majority of probe sets are down-regulated by PPAR $\gamma$	<b>121</b>
<b>Figure 20:</b>	HotellingT2 score is positively correlated to regulation	<b>123</b>
<b>Figure 21:</b>	Time course kinetic classes representing steady-state regulation	<b>126</b>
<b>Figure 22:</b>	Time course kinetic classes representing linear, non- steady-state regulation	<b>128</b>
<b>Figure 23:</b>	Time course kinetic classes representing regulation which can not be explained	<b>130</b>
<b>Figure 24:</b>	Gene ontology classes matching known functions of PPAR $\gamma$	<b>133</b>
<b>Figure 25:</b>	Lipid metabolism: prototypical GO term subclass of the metabolism superfamily regulated by PPAR $\gamma$	<b>135</b>
<b>Figure 26:</b>	DNA replication: prototypical GO term subclass of proliferation regulated by PPAR $\gamma$	<b>138</b>
<b>Figure 27:</b>	Morphogenesis: prototypical GO term subclass of invasion regulated by PPAR $\gamma$	<b>140</b>
<b>Figure 28:</b>	Angiogenesis and Calcium-mediated Signaling: ontology classes not expected to be regulated by PPAR	<b>142</b>
<b>Figure 29:</b>	Ingenuity Pathways Analysis Inferred DSCR1-Calciuneurin-NFATc pathway mediated by PPAR $\gamma$	<b>145</b>
<b>Figure 30:</b>	Ingenuity Pathways Knowledge Base canonical pathway of calcium-mediated signaling with PPAR $\gamma$ regulation overlay	<b>147</b>
<b>Figure 31:</b>	Functions involved in proliferation overlaid on DSCR1-Calciuneurin-NFATc inferred model	<b>150</b>
<b>Figure 32:</b>	Functions involved in invasiveness overlaid on DSCR1-Calciuneurin-NFATc inferred model	<b>152</b>
<b>Figure 33:</b>	DSCR1 isoform 1 is expressed in MOSER S cells	<b>154</b>
<b>Figure 34:</b>	DSCR1 is induced by RS5444 in MOSER S cells	<b>157</b>

<b>Figure 35:</b>	DSCR1 induction by thiazolidinediones is dose-dependent	<b>159</b>
<b>Figure 36:</b>	DSCR1 is a secondary target of PPAR $\gamma$	<b>162</b>
<b>Figure 37:</b>	Calcineurin activity is required for PPAR $\gamma$ -induced DSCR1 expression	<b>164</b>
<b>Figure 38:</b>	Characterization of DSCR1 sh-RNA lentiviral candidates	<b>167</b>
<b>Figure 39:</b>	DSCR1 sh-RNA B71 inhibits DSCR1 induction by PPAR $\gamma$	<b>169</b>
<b>Figure 40:</b>	Silencing of DSCR1 gene blocks PPAR $\gamma$ induced suppression of MOSER S cell proliferation and invasion	<b>171</b>

## LIST OF ABBREVIATIONS

15d-PGJ2	15-deoxy-D <sup>12,14</sup> -prostaglandin J <sub>2</sub>
ACF	Aberrant Crypt Foci
ACO	Acyl-CoA Oxidase
ANOVA	Analysis Of Variance
AOM	Azoxymethane
AP-1	Heterodimeric Transcription factor formed from c-Fos and c-Jun gene
APC	Adenomatosis Polyposis Coli gene
AR	Androgen Receptor gene
ARRB1	Arrestin Beta 1 gene
CCN1	Cyclin A2 gene
CEACAM6	Carcinoembryonic Antigen-Related Cell Adhesion Molecule 6 gene
CFTR	Cystic Fibrosis Conductance Regulator gene
CHX	Cyclohexamide
c-MYC	Myelocytomatosis Viral Oncogene Homolog (avian) gene
CnA	Calcineurin A sub-unit gene
CnB	Calcineurin B sub-unit gene
COUP-TF	chicken ovalbumin upstream promoter transcription factor gene
COX-2	cyclooxygenase-2 gene
cPGI	Carbaprostacyclin
D3T	3H-1,2-dithiole-3-thione (anticarcinogen)
DACH1	Dachshund homolog 1 gene
DBD	Deoxyribonucleic Acid Binding Domain
DCOM	Distributed Component Object Model
DMEM	Dulbecco/Vogt Modified Eagle's Minimal Essential Medium
DMSO	Dimethyl sulfoxide (CH <sub>3</sub> ) <sub>2</sub> SO
DSCR1	Down Syndrome Critical Region 1 gene
EC <sub>50</sub>	Effective Concentration 50 percent
EDTA	Ethylenediaminetetraacetic Acid
ER	Estrogen Receptor gene
FBS	Fetal Bovine Serum
GCOS	GeneChip® Operating Software (Affymetrix)
GC-RMA	Guanine-Cytosine Robust Microarray Analysis (Bioconductor)

GO	Gene Ontology
GR	Glucocorticoid Receptor gene
GW9662	2-Chloro-5-nitro-N-phenyl-benzamide (PPAR $\gamma$ antagonist)
HEK293	Human Embryonic Kidney Cells #293
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HNF-4	Hepatocyte Nuclear Factor gene
IPA	Ingenuity's Pathways Analysis
IPKB	Ingenuity's Pathway Knowledge Base
K-Ras	V-Ki-ras2 Kirsten Rat Sarcoma Viral Oncogene Homolog gene
LBD	Ligand Binding Domain
luc	Luciferase
LXR	Liver X Receptor gene
MAPK	Mitogen-activated Protein Kinase gene
MAS5	Microarray Analysis Software Version 5
MATLAB	Matrix Laboratory (Mathworks)
MB	Multivariate Bayesian statistic
MIT	Massachusetts Institute of Technology
MM	Mismatch (Affymetrix)
MOSER	Medium Off-Serum Cells
MOSER S	Medium Off-Serum (Sensitive to TGF- $\beta$ ) Cells
MvA	Mean versus Average (a.k.a. Bland-Altman) plot
NaCl	Sodium Chloride
NFAT	Nuclear Factor of Activated T-cells
NFATc	Nuclear Factor of Activated T-cells (cytoplasmic component)
NF $\kappa$ B	Nuclear Factor kappa B
PBS	Phosphate-Buffered Saline
PCR	Polymerase Chain Reaction
PDK4	Pyruvate Dehydrogenase Kinase 4 gene
PM	Perfect Match (Affymetrix)
PNN	Pinin, Desmosome Associated Protein gene
PPAR	Peroxisome Proliferator-Activated Receptor nuclear hormone superfamily
PPAR $\gamma$	Peroxisome Proliferator-Activated Receptor Gamma Isoform gene
PPP3CA	Calcineurin Regulatory Subunit A gene
PPP3CB	Calcineurin Regulatory Subunit B gene
PPRE	Peroxisome Proliferator-Activated Receptor Element
PR	Progesterone Receptor gene

PUFA	Polyunsaturated Fatty Acid
PVDF	Polyvinylidene Fluoride
qPCR	Quantitative Real-Time Polymerase Chain Reaction
RACE	Rapid Amplification of cDNA Ends
RAR	Retinoic Acid Receptor gene
RHOB	Ras Homolog Gene Family, Member B gene
RIE	Rat Normal Intestinal Epithelial Cells
RITR	RNAi Interference Technology Resource
RMA	Robust Microarray Analysis (bioconductor)
RXR	Retinoid X Receptor gene
RZD	Retinoid Z Receptor gene
SF-1	Steroidogenic Factor 1 gene
shRNA	Short-Hairpin Ribonucleic Acid
SYNC1	Syncoilin, Intermediate Filament 1 gene
TBS	Tris Buffered Saline
TBST	Tris Buffered Saline plus Tween20
TGF $\beta$	Transforming Growth Factor Beta gene
TIFF	Tagged Image Format File
TR	Thyroid Hormone Receptor gene
TZDs	Thiazolidinediones (PPAR $\gamma$ antagonist)
UTMB	The University of Texas Medical Branch (Galveston, Texas)
VEGF	Vascular Endothelial Growth Factor gene
VTR	Vitamin D Receptor gene

## GLOSSARY

<i>a priori</i>	<i>from first principles.</i> Derived by or designating the process of reasoning without reference to particular facts or experience
adenocarcinoma	A malignant tumor originating in glandular tissue
adenoma	A benign epithelial tumor having a glandular origin and structure
Affymetrix	The world's leading manufacturer of DNA microarrays, based in Santa Clara, California, United States
Age of Enlightenment	The 18 <sup>th</sup> -century intellectual movement <i>The Enlightenment</i> , which advocated reason as a means to establishing an authoritative system of aesthetics, ethics, government, and logic, which would allow human beings to obtain objective truth about the universe.
Age of Reason	The 17 <sup>th</sup> -century philosophy in the West is generally regarded as seeing the start of modern philosophy and is considered to succeed the Renaissance and precede the Age of Enlightenment.
agonist	A drug or other chemical that can combine with a receptor on a cell to produce a physiologic reaction typical of a naturally occurring substance
anchorage-dependent growth	Cellular growth predicated on attachment to a surface (usually plastic culture dishes)
androgen	A steroid hormone, such as testosterone or androsterone, that controls the development and maintenance of masculine characteristics
angiogenesis	The formation of new blood vessels
antagonist	A chemical substance that interferes with the physiological action of another, especially by combining with and blocking its nerve receptor
antioxidant	A chemical compound or substance that inhibits oxidation
ArrayAnalyzer	The Insightful ArrayAnalyzer web-based solution is integrated as part of S-PLUS Server, which includes S-PLUS as the statistics and graphics engine and is used to analyze genechip data.

autosomal recessive	A pattern of inheritance in which both copies of an autosomal gene must be abnormal for a genetic condition or disease to occur. An autosomal gene is a gene that is located on one of the autosomes or non-sex chromosomes. When both parents have one abnormal copy of the same gene, they have a 25% chance with each pregnancy that their offspring will have the disorder
azoxymethane	A potent carcinogen used to induce colon cancer in rats and mice
bacteriophage	Any of the viruses that infect bacterial cells with dimensions from about 20 to about 200 nanometers
Bayesian (statistics)	An approach to statistics in which estimates are based on a synthesis of a prior distribution and current sample data. Bayesian statistics is not a branch of statistics in the way that, for example, nonparametric statistics is. Bayesian statistics is a self-contained paradigm providing tools and techniques for all statistical problems. In the classical frequentist viewpoint of statistical theory, a statistical procedure is judged by averaging its performance over all possible data. However, the Bayesian approach gives prime importance to how a given procedure performs for the actual data observed in a given situation.
Benjamini & Hochberg	A particular approach to control type-I statistical error from multiple testing
Beta-catenin	A subunit of the cadherin protein complex and is a component of the canonically wnt-signalling pathway
bifurcation treatments	A change in the stability or in the types of solutions which occurs as a parameter is varied in a dissipative dynamic system. The change can be seen as a bifurcation point in a graph of the parameter being varied vs. One of the properties of the solutions
Bioconductor	A free, open source and open development software project for the analysis and comprehension of genomic data
biostatistics	Application of statistics to the analysis of biological and medical data
biotinylation	The process of adding a Biotin tag to a molecule or surface
Bonferroni	States that if an experimenter is testing $n$ independent hypotheses on a set of data, then the statistical significance level that should be used is $n$ times smaller than usual
boot-strap	(in statistics) The use of a subset of experimental data to test the performance of a model which will be used to analyze the dataset. Refers to the adage "pulling oneself up by their bootstraps."

caecal tumors	Tumors of or related in origin to the caecum (large blind pouch forming the beginning of the large intestine)
calcineurin	Dimeric serine/threonine phosphatase composed of a catalytic CnA subunit and a regulatory CnB subunit
calsupressin	A protein suppressing the activity of calcineurin proteins
carcinogen	A cancer-causing substance or agent
carcinogenesis	The processes of tumor development
Carolus Linnaeus	Swedish botanist and founder of the modern classification system for plants and animals
Celera	Company established in May 1998 by the Perkin-Elmer Corporation (and was later purchased by Applera Corporation), with Dr. J. Craig Venter from The Institute for Genomic Research (TIGR) as its first president. While at TIGR, Venter and Hamilton Smith led the first successful effort to sequence an entire organism's genome, that of the <i>Haemophilus influenzae</i> bacterium and later the human genome while at Celera.
chicken ovalbumin	The albumin (water-soluble protein) of avian egg whites
ChIP-on-Chip	A bioinformatics technique of incorporating protein binding information obtained through chromatin immunoprecipitation with information from microarrays. The data is synthesized to present a map of protein binding effects on gene transcription.
Chromas Lite	Software used to analyze chromatogram (gene sequencing) data
Clark's theory	This theory makes the assumption that there exists a linear relationship between an agonist's occupation of receptor binding sites and the downstream response
coactivators	Any of a number of proteins and other molecules classified loosely as responsible for activating gene transcription
colorectal cancer	Cancer of the colon is the disease characterized by the development of malignant cells in the lining or epithelium of the first and longest portion of the large intestine
conjugated	(chemistry) Formed by the union of two compounds
connectivity map	(statistics) A method in graph theory for visual displaying correlation strength of many variables
corepressors	Any of a number of proteins and other molecules classified loosely as responsible for repressing gene transcription

covariance matrix	In statistics and probability theory, the covariance matrix is a matrix of covariances between elements of a vector. It is the natural generalization to higher dimensions of the concept of the variance of a scalar-valued random variable.
Ct value	A metric used in analyzing quantitative real-time PCR data. The number represents the number of machine thermal cycles at a given (arbitrary) cut-off point.
cycloheximide	An inhibitor of protein biosynthesis in eukaryotic organisms produced by the bacterium <i>Streptomyces griseus</i> . Cycloheximide exerts its effect by interfering with peptidyl transferase activity of the 60S ribosome, thus blocking translational elongation.
cyclooxygenase-2	An enzyme that makes prostaglandins that cause inflammation and pain and fever
cyclophilin A	An abundant cytoplasmic protein that catalyzes cis-trans isomerizations and has high affinity for the immunosuppressive drug cyclosporin A
cyclosporine A	An immunosuppressive drug obtained from certain soil fungi, used mainly to prevent the rejection of transplanted organs by blocking the activity of the serine-threonine protein phosphatase calcineurin.
dChip	DNA-Chip analysis model-based software developed at Harvard
dephosphorylation	The inverse process of phosphorylation
dimerization	A chemical reaction in which two identical molecular entities react to form a single dimer (i.e. one functional protein consisting of two proteins)
dynamic Bayesian networks	A Bayesian network is a form of probabilistic graphical model, also known as Bayesian belief network or just belief network. A Bayesian network can be represented by a graph (as in graph theory) with probabilities attached. Thus, a Bayesian network represents a set of variables together with a joint probability distribution with explicit independence assumptions
electropherograms	A record of the results of an electrophoresis, such as a filter paper on which the components of a mixture are deposited as they migrate under the influence of an electric field.

electrophoresis	A method of separating substances, especially proteins, and analyzing molecular structure based on the rate of movement of each component in a colloidal suspension while under the influence of an electric field
electrophoretic assay	An assay or protocol of electrophoresis
endothelium	A thin layer of flat epithelial cells that lines serous cavities, lymph vessels, and blood vessels
epithelial cells	Cells that form a thin surface coating on the outside of a body structure
ermineJ	Software used to analyze gene ontologies from sets of genechip data
expert-system	A computer algorithm that uses available information, heuristics, and inference to suggest solutions to problems in a particular discipline. See machine-learning
False Detection Rate	A statistical method used in multiple hypothesis testing to correct for multiple comparisons
fenofibrate	A PPAR $\alpha$ agonist
floxed alleles	Molecular Biology in vivo technique (usually in rodents) used to make conditional knock-outs of particular genes using Cre-Lox technology. The inactivation of the gene is determined by the progeny of mating.
F-statistic	Value calculated by the ratio of two sample variances. The F statistic can test the null hypothesis: (1) that the two sample variances are from normal populations with a common variance; (2) that two population means are equal; (3) that no connection exists between the dependent variable and all or some of the independent variables
functional genomics	A field of molecular biology that attempts to make use of the vast wealth of data produced by genomic projects (such as genome sequencing projects) to describe gene and in some cases protein functions and interactions
gaussian models	Statistical model assuming normal distributions
geimsa	A complex of stains specific for the phosphate groups of DNA, specifically regions having high A-T bonding.

gene ontology	The Gene Ontology, or GO, project can be broadly split into two parts. The first is the ontology itself, a controlled vocabulary of terms split into three related ontologies covering basic areas of Molecular biology: the molecular function of gene products, their role in multi-step biological processes, and their localization to cellular components. The ontology is continuously updated, and new versions are made available on a monthly basis. The second part is annotation, the characterization of gene products using terms from the ontology. The members of the GO Consortium submit their data and it is made publicly available through the GO website
GeneChip®	Affymetrix term for a DNA Microarray: A wafer of photolithographic applied grids of spots containing identical single-stranded polymeric molecules of deoxyribonucleotides, usually oligonucleotides or complementary DNAs, attached to a solid support (such as a membrane, a polymer, or glass) used to simultaneously analyze the expression levels of the corresponding genes
genomics	The study of all of the nucleotide sequences, including structural genes, regulatory sequences, and noncoding DNA segments, in the chromosomes of an organism
gentamycin	A broad-spectrum antibiotic derived from an actinomycete of the genus <i>Micromonospora</i> , used in its sulfate form to treat various infections. Used extensively in cell culture to control infection
glucocorticoid	Naturally-produced steroid hormones, or synthetic compounds, that inhibit the process of inflammation
heterodimer	A dimer of two dissimilar bound proteins
hexanucleotide	A nucleotide sequence containing six nucleic acids.
HGU-95Av2	Affmetrix designation of a genechip used to measure the human genome
high-throughput	(bioinformatics) Any experimental technique which can measure a great number of variables at once (e.g. microarray technology)
homodimer	A dimer of two similar or identical proteins
HotellingT2	A test and associated figure of merit for the Hotelling's T2 distribution
hybridization	Production of a hybrid by pairing complementary ribonucleic acid and deoxyribonucleic acid (DNA) strands
hyperparameter	A single parameter incorporating a model with many variables

immunoprecipitation	A protein purification method which involves the formation of an antibody-protein complex to separate out the protein of interest
<i>in silico</i>	An expression used to mean performed on computer or via computer simulation
Ingenuity	A software company who produces, markets, and distributes the web-based systems biology software package Ingenuity Pathways Analysis
Insightful	A software company who produces and markets the statistical software package S-Plus.
integrin	A heterodimeric transmembrane receptor protein of animal cells that binds to components of the extracellular matrix on the outside of a cell and to the cytoskeleton on the inside of the cell, functionally connecting the cell interior to its exterior; in blood cells, integrins are also involved in cell-cell adhesion
interpolate	To estimate a value(s) of a function or series between two known values
law of mass action	Fundamental law of chemical kinetics (the study of rates of chemical reactions), formulated in 1864–79 by the Norwegian scientists Drs. Cato M. Guldberg and Peter Waage. The law states that the reaction rate of any simple chemical reaction is proportional to the product of the molar concentrations of the reacting substances, each raised to the power corresponding to the number of molecules of that substance in the reaction.
lentivirus	A genus of slow viruses of the <i>Retroviridae</i> family, characterized by a long incubation period. Lentiviruses can deliver a significant amount of genetic information into the DNA of the host cell, so they are one of the most efficient methods of a gene delivery vector. HIV, SIV, and FIV are all examples of lentiviruses
longitudinal	Concerned with the development, analysis and/or study of data of persons or groups over time
lowess	or Loess: a more robust and modern version of the classical linear and nonlinear least squares regression
machine learning	The process or technique by which a device modifies its own behavior as the result of its past experience and performance (e.g. learn through reinforcement)
MATLAB	A powerful linear algebra software package developed by Mathworks used for scientific and technical computing

matrigel	The trade name for a gelatinous protein mixture secreted by mouse tumor cells and marketed by BD Biosciences. This mixture resembles the complex extracellular environment found in many tissues and is used by cell biologists as a substrate for cell culture.
MaxDiff	Also, MaxDiff_of_Medians: The maximum difference of median values within a kinetic profile between one treatment condition and another treatment condition
MB-statistic	A statistical metric created by Tai and Speed at U.C. Berkeley which is similar to the Bayesian log-odds ratio and is used as a measurement of confidence.
medianpolish	Invented by the American statistician Dr. John Tukey in 1977. This is the summarization used in the RMA expression summary algorithm. A multichip linear model is fit to data robustly and results in log <sub>2</sub> scale values representing the transcript abundance of a biological sample.
mesenchymal stem cells	(MSC) Multipotent stem cells that can differentiate into a variety of cell types. Cell types that MSCs have been shown to differentiate into in vitro or in vivo include osteoblasts, chondrocytes, myocytes, adipocytes, neuronal cells, and, as described lately, into beta-pancreatic islets cells.
metabonomics	The quantitative measurement of the dynamic multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modification. This approach originated in Imperial College London in Dr. Jeremy Nicholson's laboratory and has been used in toxicology, disease diagnosis and has expanded to a number of other fields
metastatic lesions	A cancer that has spread to an organ or tissue from a primary cancer located elsewhere in the body
microarrays	see genechip
moderation	Employing assumptions into a statistical model which analyzes genechip data to mitigate wild fluctuations and resultant statistical error
monolayers	Cell culture limited to growth in one plane on the surface of a culture dish (that is not growing on top of one another).
monotonic	Always increasing or always decreasing, as the value of the independent variable increases; of a function
multivariant	Having or involving more than one variable: <i>multivariate statistical analysis</i>
myocyte	A contractile cell. A muscle cell

myofiber	Muscle fiber: A cylindrical, multinucleate cell composed of numerous myofibrils that contracts when stimulated
neovascularization	Proliferation of blood vessels in tissue not normally containing them
non-parametric	Non-parametric models differ from parametric models in that the model structure is not specified a priori but is instead determined from data. The term nonparametric is not meant to imply that such models completely lack parameters but that the number and nature of the parameters are flexible and not fixed in advance. Nonparametric models are therefore also called distribution free
normalization	The process of removing statistical error in repeated measured data
nuclear hormone receptor	Any of a superfamily of proteins that directly regulate transcription in response to hormones and other ligands
oligonucleotide	A deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) sequence composed of two or more covalently linked nucleotides
oncogeneticist	A general term of a biologist principally researching the cause of cancer and its underlying genetic and genomic mechanisms of action
oncogenic	Tending to cause or give rise to tumors
oncogenic mutations	Mutations in DNA giving rise to the cause of tumors
orphan receptors	The class of nuclear receptors which have been identified, but whose ligand have yet to be elucidated
osteoblast	A cell from which bone develops; a bone-forming cell
over-representation	Determining the statistical significance of a particular gene-ontology class by calculating its relative over-representation compared to other classes within a predefined set of genes
peroxisome	A cell organelle containing enzymes, such as catalase and oxidase, that catalyze the production and breakdown of hydrogen peroxide
phase portrait	A geometric representation of the trajectories of a dynamical system in the phase plane
phenotype	The observable physical or biochemical characteristics of an organism, as determined by both genetic makeup and environmental influences
phosphatase	Any of numerous enzymes that catalyze the hydrolysis of esters of phosphoric acid and are important in the absorption and metabolism of carbohydrates, nucleotides, and phospholipids and in the calcification of bone

phospholipid	Any of various phosphorous-containing lipids, such as lecithin and cephalin, which are composed mainly of fatty acids, a phosphate group, and a simple organic molecule. Also called <i>phosphatide</i>
phosphorylation	The esterification of compounds with phosphoric acid. Also an essential process triggering the biological activity of biomolecules through interaction with kinases
pioglitazone	A drug that reduces the amount of glucose (sugar) in the blood. It is in a class of anti-diabetic drugs called "thiazolidinediones" that are used in the treatment of type II diabetes
polybrene	A cationic polymer used to increase the efficiency of infection of certain cells with a retrovirus in cell culture
polynomial regression	Statistical regression of data modeled by the form $a_0x^n+a_1x^{n-1}+a_2x^{n-2}+\dots+a_{n-1}x+a_n$ where $n$ is a positive integer and $a_0, a_1, a_2, \dots, a_n$ are any numbers.
posterior probability	The probability of a hypothesis being true conditioned on certain data
post-mitotic	Transformed cells which have stopped dividing
PPAR $\gamma$	The gamma isoform of Peroxisome proliferator-activated receptors (which in turn belongs to the superfamily of nuclear hormone receptors) that is activated by fatty acids and thiazolidinedione drugs and plays a role in insulin sensitivity and adipogenesis.
progesterone	A steroid hormone, $C_{21}H_{30}O_2$ , secreted by the corpus luteum of the ovary and by the placenta, that acts to prepare the uterus for implantation of the fertilized ovum, to maintain pregnancy, and to promote development of the mammary glands
prostaglandins	Any of a group of potent hormone-like substances that are produced in various mammalian tissues, are derived from arachidonic acid, and mediate a wide range of physiological functions, such as control of blood pressure, contraction of smooth muscle, and modulation of inflammation
proteomics	Refers to the study of all the proteins expressed by a genome; proteomics involves the identification of proteins in the body and the determination of their role in physiological and pathophysiological functions
pseudo-code	In software engineering, an outline of a program written in the user's natural language (e.g. English) and used to plan the program or communicate the program's statement of work at a sufficient level of abstraction to make clear its intended purpose
quantile	A value which divides a set of data into equal proportions

quantile-quantile plot	Q-Q plot: a tool for diagnosing differences in distributions (such as non-normality) of a population from which a random sample has been taken
R language	A open source programming language and software environment for statistical computing and graphics; an implementation of the statistical language S
recursive	The ability of a subroutine, program module or statement of work to call itself. It is helpful for writing routines that solve problems by repeatedly processing the output of the same process
reductionism	An attempt or tendency to explain a complex set of facts, entities, phenomena, or structures by another, simpler set
René Descartes	Often called the father of modern science. He established a new, clear way of thinking about philosophy and science by rejecting all ideas based on assumptions or emotional beliefs and accepting only those ideas which could be proved by or systematically deduced from direct observation
reverse-engineering	The analysis of a completed system in order to isolate and identify its individual components or building blocks
ribosomal RNA	Non-coding RNA that is a permanent structural part of the ribosome
Richard Dawkins	a British ethologist, evolutionary scientist, and popular science writer who holds the Charles Simonyi Chair in the Public Understanding of Science at Oxford University
rosiglitazone	A drug that reduces the amount of glucose (sugar) in the blood. It is in a class of anti-diabetic drugs called thiazolidinediones that are used in the treatment of type II diabetes
RS5444	A high potency thiazolidinedione manufactured by Sankyo Pharmaceuticals, Ltd.
<i>Saccharomyces cerevisiae</i>	a.k.a Baker's or Brewer's yeast is a species of budding yeast heavily used in the field of genetics and functional genomics because it is one of the few eukaryotes whose entire genome has been annotated
spline	A smoothing curve that runs through a series of given points. Also a piecewise polynomial function that can have a locally very simple form, yet at the same time be globally flexible and smooth.
steroidogenic factor	A member of the sub-family number 5 of nuclear hormone receptors

streptomycin	An antibiotic, $C_{21}H_{39}O_{12}N_7$ , produced by the actinomycete <i>Streptomyces griseus</i> , used to treat tuberculosis and other bacterial infections
Stuart Kauffman	A theoretical biologist and complex systems researcher who is best known for arguing that the complexity of biological systems and organisms might result as much from self-organization and far-from-equilibrium dynamics as from Darwinian natural selection
summarization	In the analysis of genechips, the statistical step giving a numerical value the abundance of transcript in a biological sample
superfamily	The general class of a particular group of proteins. See nuclear hormone receptors
Systems Biology	The study of the interactions between the components of a biological system, and how these interactions give rise to the function and behavior of that system (for example, the enzymes and metabolites in a metabolic pathway) through the use of computation and informatics and relies heavily on a holistic approach to solving biological problems, rather than the traditional reductionist approach of molecular biology. The holistic approach acknowledges that most, if not all, biological problems are essentially dynamic non-linear events and are therefore incapable of being understood completely without sophisticated computational modeling of hyper-dimensional problem spaces through time
thiazolidinediones	(TZDs) A class of compounds that act by binding to PPARs (peroxisome proliferator-activated receptors), a group of receptor molecules inside the cell nucleus, specifically PPAR $\gamma$ . The normal ligands for these receptors are free fatty acids (FFAs) and eicosanoids. When activated, transcription of specific target genes is initiated
transactivation	Activation of a gene or protein indirectly
transcription	The biological process in which the DNA sequence of a gene is copied into mRNA; the process whereby a base sequence of messenger RNA is synthesized on a template of complementary DNA
transfection	The introduction of foreign DNA into eukaryotic or prokaryotic cells, such as animal or bacterial cells. Transfection typically involves opening transient holes or gates in cells to allow the entry of extracellular molecules, typically supercoiled plasmid DNA, or siRNA, among others

transgene	A segment of DNA containing a gene sequence which has been isolated from one organism and is introduced into a different organism
transwell chamber	Also called a Boyden chamber: In biology, an instrument using a series of wells and membranes used to measure various properties of a cell; typically motility and invasiveness.
troglitazone	An oral antidiabetic agent from the thiazolidinedione drug class that lowers blood glucose by improving the response of target cells to insulin without increasing insulin secretion
trypsin	A pancreatic enzyme that catalyzes the hydrolysis of proteins to form smaller polypeptide units, thereby degrading proteins (proteinase).
trypsinization	Using the proteinase trypsin to detach cell culture from the culture plate for experimental manipulation
Tukey's biweight	Statistical function used in estimation of robustness
tumorigenesis	Formation or production of tumors
type-I statistical error	The error of rejecting a null hypothesis when it is the true state of nature. In other words, this is the error of accepting an alternative hypothesis (the real hypothesis of interest) when an observation is due to chance
type-II statistical error	The error of not rejecting a null hypothesis when the alternative hypothesis is the true state of nature. In other words, this is the error of failing to accept an alternative hypothesis when you don't have adequate statistical power
unsupervised clustering	A method of machine learning where a model is fit to observations. In the case of clustering, the model's goal is to separate data into recognizable sets or classifications. It is distinguished from supervised learning by the fact that there is no <i>a priori</i> output. In unsupervised learning, a data set of input objects is gathered. Unsupervised learning then typically treats input objects as a set of random variables. A joint density model is then built for the data set
vitalism	The theory or doctrine that life processes arise from or contain a nonmaterial vital principle and cannot be explained entirely as physical and chemical phenomena
webcrawlers	A computer program that searches for information on the world wide web and is usually coded utilizing principles in the fields of artificial intelligence and statistical machine learning.

William of Occam

A 14th century English philosopher and Franciscan friar. Resistant to the popular wave of Scholasticism, a philosophical position that tried to unify worldly and religious ideas, William of Ockham asserted that one could not know God through reason and rationality. His philosophy is sometimes called nominalism, and he is now most famous for only one of his many ideas, what is called the principle of Ockham's Razor (or The Law of Parsimony)

xenobiotic  
xenopus

Foreign to the body or to living organisms  
Any of various aquatic, tongueless, clawed frogs of the genus *Xenopus*, native to southern Africa, especially *X. laevis*, widely used in the study of vertebrate development because their exceptional large eggs provide ease of experimental observation and manipulation

# CHAPTER 1: BACKGROUND

## SYSTEMS BIOLOGY

### The Systems Biology Paradigm Shift

The desire to find order within biological systems has held scientific interest since the days of Aristotle. In the mid-eighteenth century, the Swedish scholar Carolus Linnaeus used similarities among known organisms to classify them into his *Systema Naturae* (Natural Systems), a method still used in biology today (1). Explanations for the natural order of life span scientific thought from Darwin's *Origin of Species* to Stuart Kauffman's "order-for-free" theory (2, 3). The hidden subtleties of nature's complexity can be expressed by a simple question that biologists continue to struggle with: "If Darwin is correct, then how could something as complex as the eye have evolved from the primordial soup?" Kauffman ventured in his seminal book *The Origins of Order* that spontaneous, emergent behavior is a possible explanation to this apparent paradox (4). He was very careful in promoting his theory that self-organization is a natural property of complex genetic systems, for self-organization and emergent systems hint of vitalism, a once popular but now discredited notion that much of the wonder of the natural world is the consequence of an *élan vital* (vital spirit). Nature wasn't so much explained as explained away by this notion (3).

In light of increased skepticism from the scientific community, biology moved away from the metaphysical realm of vitalism towards reductionism *a priori* (from first

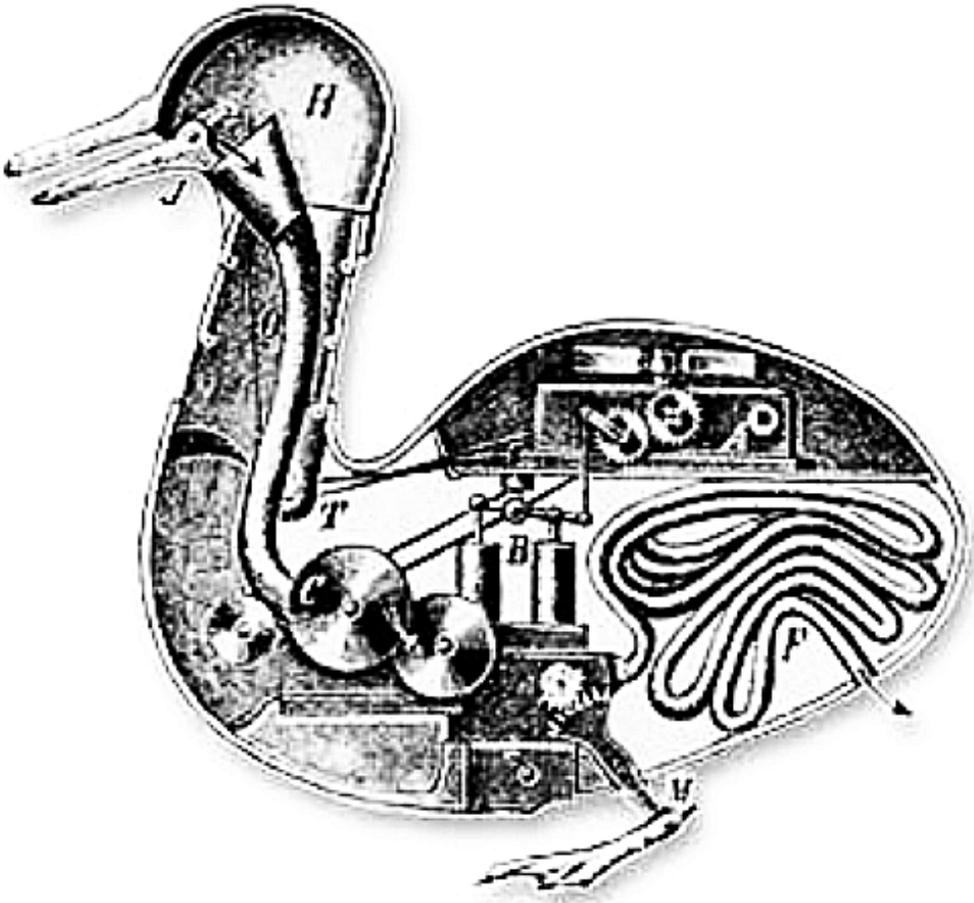
principles). Reductionism, an approach that today drives much of modern biology, came into fashion during the Age of Reason and the Age of Enlightenment. This philosophy theorizes that the nature of any complex system can be reduced and therefore explained by a series of smaller and/or simpler elements. One of the first champions of reductionism was the philosopher and mathematician René Descartes, who maintained that unlike human beings who were considered divine creation, animals could be reductively explained as automata. This idea was famously captured in a freehand drawing depicting the internal workings of a duck as mechanics in part V of Descartes' thesis Discourses, 1637 [Figure 1] (5). Descartes argued that because organisms could be thought of as machines, they could be taken apart, examined piece by piece, and then put back together to understand each part in the context of the whole.

Our appreciation for the simple, yet elegant concept of reductionism has evolved along with our drive to understand more. Historically, this appreciation has proved to be a matter of scale as technology pushed science into the microscopic and molecular world. Richard Dawkins in his book *The Selfish Gene* continued the reductionist tradition into the molecular world by arguing that if genes were the fundamental components of life, then all life and all natural behavior could be integrated back to the study of genetic mechanisms (6).

This born-again reductionist fashion in life-science research continued, bound tightly to hypothesis driven research. Experimental design was straightforward. Biological systems were deconstructed to the molecular level and schematically categorized using molecular biology and biochemical techniques. This approach served

**Figure 1: *Illustration of automata in Descartes' thesis Discourses, 1637***

Réne Descartes maintained that much of life could be reduced mechanistically as automata. This view was part of a much wider held philosophy of reductionism, in which the nature of any complex system can be reduced into subsets of smaller, more fundamental sets. Descartes reasoned that the animals, in this case a duck, could be explained as a series of mechanical parts. [Modified from (5)]



the biological community well for over sixty years beginning in earnest with Beadle and Tatum demonstrating a precise relationship between genes and proteins (7). However, since the completion of the first rough draft of the human genome (8) and the resultant technologies that emerged from this work, biologists have been naturally pushed toward a new view of biology—a more holistic approach known as Systems Biology. The maturing of thought from reductionism to holistic systems biology is reflected in the progress of Dawkins’ writings. In 1986, Dawkins introduced the term “hierarchical reductionism” in his book *The Blind Watchmaker* (9). The concept of reductionism was thereby expanded to accommodate the hypothesis that while any complicated mechanism or system can be broken down into smaller component parts, the hierarchal division of work may and probably will lose context as one progress further down the spiral of simpler and less complicated internal workings. This is not an entirely new idea. William of Occam, the Franciscan friar best known for his Razor, offered this theorem of complexity in the 14<sup>th</sup> century. The Razor in Latin was articulated in various forms including: “*entia non sunt multiplicanda praeter necessitatem*” which literally translates into “*entities should not be multiplied beyond necessity.*” Others, following Bartlett’s Quotations understand Occam’s Razor as “*pluralites non est ponenda sine necessitate*”, which is interpreted as “*complexity does not exist for no reason.*” (10) (As a side note, the Razor is often misinterpreted to mean “*all things being equal, the simplest explanation is the most probable.*”) Regardless of the precise quotation, Dawkin’s concept of hierarchal reduction touches very close to the Razor and acknowledges the insight that early scientific thinkers had for complex systems. Moreover, it is clear that

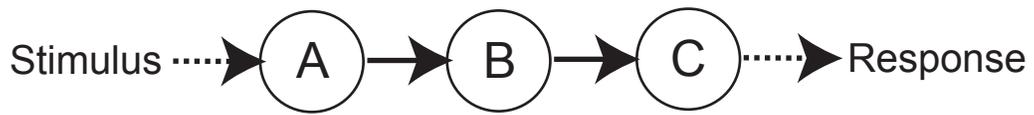
scholarly insight into complexity points to a new way of thinking about biological systems, which involves more than viewing the whole as the sum of its components. This principle is simply illustrated by considering the fundamental issues that arise when one contrasts the features of serial versus parallel pathways in biology.

Consider for a moment the simplest kind of pathway in biology where a stimulus translates through serial connections of processes into some form of reaction. A model for this is shown in **Figure 2A**, where the binary interactions between any two processes form a chain. In the simplest pathway, these processes would represent one kind of molecular species. A stimulus causes molecule A to interact with molecule B, molecule B then interacts with molecule C and so forth until the terminal reaction is achieved. Studying these kinds of pathways is straightforward. The experimenter stimulates the system and records the resultant reactions. The model provides a means to track the progression of the stimulus, through a series of dependent connections, to the effect. And of course, the experiment can be manipulated in several ways to study any individual component of the serial connection. This reductionist approach has proven extremely successful in experimental biology. As a result, there is an immense amount of mature technologies and methods available to study binary interactions. What we know most about modern molecular biology has been learned from this approach. The drawback to this approach, of course, is that signal transduction is seldom binary. Signal transduction tends to be highly parallel, non-linear, and with redundant systems of connections. And as a result, a single stimulus may start two or more signaling pathways that influence the outcome [**Figure 2B**]. In the extreme case, a stimulus can cause an array of redundant

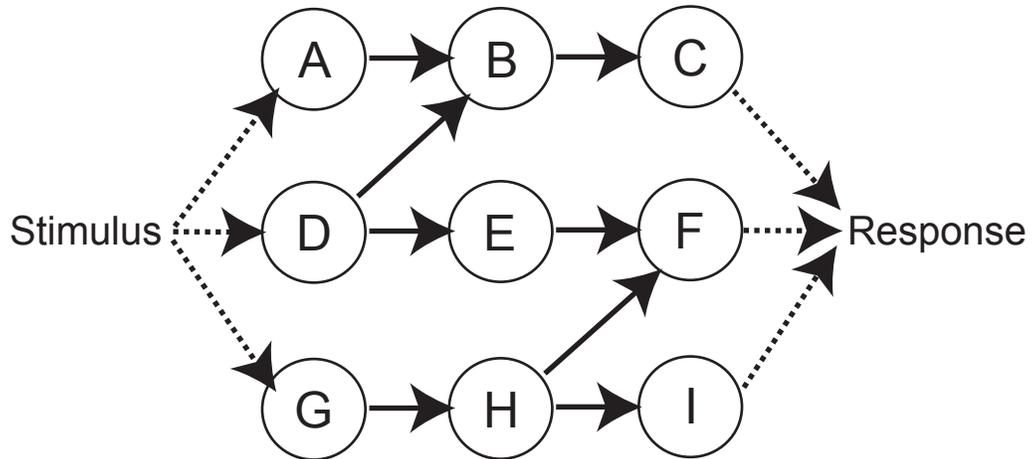
**Figure 2: *Serial versus parallel signal transduction***

**A)** A schematic modeling binary signal transduction in series. Here, the information of a stimulus is passed through biological nodes, one at a time, and ultimately manifests as a reaction. **B)** The serial model is modified to allow multiple independent and parallel pathways to pass information from a stimulus to a reaction.

A.



B.



pathways that transfer information between each other in often elaborate condition-specific ways. Highly parallel connections are now generally thought to be the rule instead of the exception which presents an obvious problem to conventional experimental technique. It is difficult to imagine how even a simple parallel pathway, such as that illustrated in **Figure 2**, could be understood in detail by analysis of a series of binary interactions, and the problem becomes more complex when one attempts to accommodate feedback and feed forward loops in the pathway. The inherent complexity of highly parallel pathways is now a prime motivator for more research in systems biology techniques.

The current systems biology paradigm requires the combination of techniques to define a representative model of the biological question. For example, global measurement techniques like oligonucleotide microarrays and 2-D gel proteomics define a system-wide high-level abstraction of the biological picture. The low-level details of this picture are filled in with more traditional molecular biology and biochemistry techniques. What follows is a modern example of the current framework. One systems-level method for understanding a biological question is to begin by defining and characterizing a network pathway (11). The first step of this method is to break the system down into components, much as a molecular biologist would approach the problem, and then categorically define the parts. Likewise, it is also important to define the physical and chemical interactions modulating molecular and cellular structure. The next step is to define the system's dynamics, or how the system behaves, under various experimental conditions. Such tests may include metabolic analysis, sensitivity analysis,

and dynamic analysis like phase portrait and bifurcation treatments (11). Next, a method must be designed that will control key elements, “nodes,” of the network that ultimately control the cell’s state. As different controls are implemented, global analysis methods are employed to capture downstream information as it propagates from one node to the next through the network. Finally, the large amounts of data are collected and archived in some form of database. These databases can then be mined by expert-system computer programs, which help the researcher form new hypotheses and theories to test.

A systems-level understanding also requires a change of perception. While molecular biology has given us insight into the individual functions of genes and proteins, the new focus is on understanding a system’s structural and dynamic context. It is much like how a pilot views flying. The pilot thinks about using the plane’s avionics system to stay on course and the plane’s control systems and surfaces, i.e. rudder, elevator, and ailerons, to keep the craft in the air. A reductionist molecular biology view of the aircraft would be more like that of the plane’s mechanic, who is more mindful of the fuselage and engine components and how they interact with one another. Granted, schematic representation of an airplane is important to the mechanic, just as schematic representation of a biological pathway is important to a molecular biologist. In fact, understanding complex systems must start with schematic representation. However, a blueprint of the plane is very limited in its ability to determine handling characteristics when something goes wrong with the control systems. Only real-time flight test data will describe characteristics of the failure. This is the aeronautical engineering equivalent to the familiar biologic methodology: the essence of discovery research.

## **Functional Genomics' Role in Discovery Research**

When Craig Venter's company Celera helped solve the first rough draft of the human genome, it began a new era discovery science in the so-called post-genomic era. Because functional genomics, that is the use of genomic projects to understand the dynamics of gene function and interaction, was the first truly reliable high-throughput technology available to biological research (12), it is therefore the most mature of several technologies along the same principle including proteomics (13, 14) (the measurement and study of large numbers of gene products, i.e. proteins) and metabonomics (15-17) (the measurement and study of large numbers of protein products). Early adoption of functional genomics as a starting point for more sophisticated research projects has been discussed extensively elsewhere (18-21) and is now a common and excepted practice (22-26).

The discovery science approach takes over where reductionist biology left off, a move from categorizing to a more pro-active feed-forward schema. Essentially the cycle is this: data create experimental opportunity, which create data, which create opportunity, and so on. The point is to let the data lay the path for the next experiment. Systems biology involves the merging of hypothesis driven research and discovery-based research with emphasis on the advancements in information technology, high-throughput data gathering techniques, biological automation techniques, and global experimental techniques.

With the wide adoption of microarrays, it has become increasingly clear that complex cell functions or phenotypes are rarely determined by a single gene. That is to say, genes rarely act individually. The advent of microarray technology has all but confirmed with overwhelming evidence that this is the case. There are exceptions such as autosomal recessive disorders like cystic fibrosis and sickle cell anemia. In cystic fibrosis a single mutation on both alleles of the cystic fibrosis conductance regulator (CFTR) causes this painful and ultimately fatal disease (27). For sickle cell anemia, a point mutation from adenine to thymine in the  $\beta$ -globin gene, results in the residue substitution from glutamic acid to valine in position number six (28). Most pathology, however, is more likely the result of many interacting genes and gene pathways working in a complex entangled web of influence. Carcinogenesis is a prime example, in which the ultimate phenotype of the tumor cell is certain to be determined by complex interactions resulting from multiple oncogenic mutations. Oligonucleotide microarrays routinely show differential genomic expression between normal cells and tumor cells a matter concerning hundreds, if not, thousands of genes. Perhaps an oncogeneticist needs only to observe a few examples of massive gene crosstalk to appreciate Stuart Kauffman's prediction of complex interplay as evidence of emergent behavior.

If Stuart Kauffman's prediction is correct, the question then becomes: "How may one consolidate system-wide data into a coherent set of hypotheses, given biology's extraordinary level of complexity?" It appears that the once esoteric branches of artificial intelligence in computer science and statistical machine learning in applied statistics may hold some clues (29, 30). For nearly three decades both disciplines worked on the same

premise that, given enough empirical data, a computer algorithm could *learn* the data's structure at some level of abstraction (31-33). As will be discussed in more detail shortly, both disciplines began to converge on a methodology and framework for predicting, displaying, and interpreting complexity utilizing techniques that have grown from initial research in artificial intelligence (34-36). These techniques have become widely appreciated in many fields of science because of advancements in affordable desktop computing power, advancements in informatics, and various technologies augmenting the internet.

### **Systems biology and reverse-engineering complexity**

The systems biology paradigm, therefore, is chiefly a means to reverse engineer a complicated system by inferring component interactions and rules given biological observation based on well planned experiments. Thus, the statement of work is to define the system first in the context of tissues or cells, then the information pathways, then the protein interactions, the proteins, the mRNA making the proteins, and finally the genes. This work flow is merely the hierarchal information flow stated above in reverse. The most generalized steps to achieve this goal follow a recursive recipe:

- 1) Define the system and its context
- 2) Discover testable hypotheses from very complicated systems through methodologies involving global measurement
- 3) Formalize this set of predictions into a cogent visualization

- 4) Incorporate previous knowledge, data and insight obtained from past experiments
- 5) Define the system at some lower level of abstraction the next iteration of experiments to perform to verify the predictions.

Eventually the systems biology community hopes mathematical models will not only provide new questions to answer, but irrefutably answer posed questions. For now, though, the community will have to settle for the recursive recipe of finding new questions to answer, while contextualizing the boundaries and rules the system operates within. Despite this shortcoming, the current systems biology paradigm does allow for rather sophisticated tools to study biological questions. Certainly some of the more advanced tools in the systems biologist arsenal are statistical machine learning algorithms to process complicated biological datasets. When these algorithms are used in the correct context, under a well defined set of assumptions, they provide a powerful vehicle for discovery research.

### **Examples of tools that have been used to solve systems problems**

#### ***Functional Genomics***

Functional genomics aims to harness the vast collection of knowledge obtained from the many genome sequencing projects around the globe. Unlike genomics and proteomics, functional genomics focuses on the dynamic aspects of molecular biology such as transcriptional control, downstream signaling, and feedback mechanisms.

Perhaps no where else on earth is this problem more appreciated than at the Broad Institute in Cambridge Massachusetts. The Broad Institute, formerly known as the Whitehead Institute, is a collaboration between the Massachusetts Institute of Technology (MIT) and Harvard University. A recent article in Science illustrates the depth of commitment the Broad holds for both functional genomics and systems biology. Todd Golub's team has produced a connectivity map which aims to connect the functions of disease to genetic perturbations and drug actions (37). In essence, the paper demonstrates the feasibility of a much larger project which databases the genomic expression of an organism under the influence of a drug action or disease state. The mapping of genomics and perturbations can be used to data mine connections between small signaling molecules sharing mechanisms of action, physiological processes, and drug mechanisms. Basically, the connectivity mapping is the beginning of a large scale community project to harness the collective wisdom of the research community and implement a systems biology mapping of cause and effect. This work is the next iteration in a string of seminal bodies of functional genomic research, from early attempts at pathway discovery (38), to disease progression (39), to early ontological analysis (40).

### ***Statistical Machine Learning for Pathway Inference***

With the explosion of data from high-throughput techniques in the life sciences comes the need for analytical methods equal to the task. Indeed, the post-genomic era has been a boon to the applied statistical community, principally among those statisticians eager to invent new bioinformatics methods capable of digesting the

enormous complexity associated with nature. Several books have recently been published on the subject (41).

The holy grail of scientific endeavors for this group of researchers is *a priori* biological network inference, where the goal is to use experimental data to calculate from first principles causal biological networks. It is the hope that these inferred networks can then be used as models for experimentation of entire systems. For example, it would be immensely useful for cancer biologists to have a working transcriptome network *in silico* of a particular class of cancer. If such a computer model of cancer existed, it is then conceivable that dry experiments could be performed involving, for example, the activation of a nuclear hormone transcription factor where statistical predictions of downstream genetic events could be calculated. Such models have been envisioned, and most share certain key elements. The first element is a model of biological high-throughput data. The now classic example is microarray data. The more sophisticated models aspire to bring divergent classes of such data together, such as microarray data in conjunction with high-throughput proteomics (42-45), as well as metabolomic data sources (15, 17).

And yet, the very first attempts to utilize principles of artificial intelligence into biological data analysis proved extraordinarily sophisticated for their time. For example, ChIP-on-Chip technology utilized chromatin immuno-precipitation in conjunction with yeast microarrays to identify the entire *Saccharomyces cerevisiae* transcriptome along with a library of basic regulatory network motifs (46-48). As genechip technology advanced and competition drove down costs, researchers began considering the possibility of incorporating temporal information into their experiments. Within the last

several years computer scientists and statisticians began to experiment, perhaps a little too eagerly, with the idea of dynamic Bayesian networks as a machine learning approach to inferring genetic pathways (49-51) from temporal expression data. Recently, individuals have employed more generalized Bayesian inference algorithms such as graphical Gaussian models with some success (52).

### ***Ingenuity Pathway's Knowledge Base (IPKB)***

Ingenuity's Pathway Knowledge Base (IPKB), the database and search engine which drives the Ingenuity Pathway's Analysis (IPA) application, first began by hiring Ph.D. level researchers as consultants to manually comb peer-reviewed journals and obtain information to later be entered into the database (Ingenuity® Systems, [www.ingenuity.com](http://www.ingenuity.com)) (53). For example, if a consultant found evidence that Gene A made protein A and protein A bind to the promoter region of gene B as a transcription factor, the interrelationship between A and B was offered to Ingenuity as legitimate data to be stored and referenced. This process of manually searching and entering information went on for some time. Eventually, Ingenuity used software robots, webcrawlers and expert systems to comb the literature automatically; ensuring accuracy of the findings with a quality control team. What has emerged from this effort is a unique corpus of information for the life sciences research community containing the largest expertly curated database of molecular biology in the world. Features of this product include: broad genome wide coverage of over 23,900 mammalian genes (10,300 human, 5,200 rat and 8,400 mouse); millions of pathway interactions extracted from literature;

semantic/linguistic consistency based on a comprehensive ontology with over 500,000 concepts; and systematic capture of canonical pathway relationships.

The creative ways in which Ingenuity Pathway's Knowledge Base has been employed by researchers to solve systems problems is remarkable. Its use as a tool spans nearly every facet of biological research from stem cell research (54, 55), to ophthalmology (56), to cancer research (37, 57-62), to proteomics (63-65), to the fledgling field of bioinformatic multivariant data integration (65-67).

Recent feature upgrades within the Ingenuity Pathway's Analysis have helped make bridging gene connections within the context of a given system particularly useful. For example, it is now possible to link two seemingly unrelated genes or proteins within the context of a biological experiment using Ingenuity Pathway's Analysis grow function. Liu and colleagues used this approach to help build a proteomic reference set of markers for trauma patients (63). Likewise, one can now classify markers within the context of canonical signaling as Si *et al.* demonstrated with CCN1/Cry61 in Wnt osteoblast differentiation of mesenchymal stem cells. Moreover, visual maps of transcriptional networks within the context of a stimulus may be constructed as Huang and colleagues did with the anticarcinogen 3H-1,2-dithiole-3-thione (D3T: an antioxidant dithiolethione used to study detoxification of environmental carcinogens) in rat cells. This approach uncovered lipid metabolism and anti-inflammatory/immune-suppressive response pathways, indicating a broader cytoprotective effect of D3T than previously expected (68).

### ***Robust Microarray Analysis (RMA)***

The underlying algorithms and assumptions employed in microarray analysis leave much room and confusion when considering a reliable approach to determine differentially expressed genes. Affymetrix provided an off-the-shelf software suite to calculate expression values and confidence for probe sets. However, recent work in the biostatistics community has demonstrated that Affymetrix's approach is flawed in its underlying assumptions. The resultant output may contain unwieldy variance in concordance with unchecked type-I error (false positives). The Robust Microarray Analysis (RMA) method first describe by Irizarry *et al.* (69) is generally considered the leading candidate in microarray analysis by the bioinformatics community based on its performance against other publicly available methods like Li and Wong's dChip (70) and Affmetrix MAS5.0 software. Irizarry *et al.* demonstrates performance gains using quantile-quantile plots of Affymetrix's spiked-in dataset, where ten positive controls and one negative control were "spiked" into a cDNA preparation for hybridization to HGU-95Av2 microarray plates (69). The reasons behind these gains are subtle and require a general understanding of how low-level microarray output is calculated.

Microarray preprocessing, that is the underlying set of steps required to provide intensities representing transcript abundance, is broken down into several steps. The first step is to correct for background noise. Affymetrix's MAS5 breaks down the microarray chip into sixteen regions and uses the lowest 2% of probe intensities per region to correct using both perfect match (PM) and mismatch (MM) probe pair information. RMA assumes a global model across the entire array and likewise assumes PM probes are

modeled as the sum of a normal distribution component and an exponential signal component. In RMA, the MM probes are not corrected and, in fact, not used at all. The prevailing wisdom holds that RMA background correction outperforms other methods (69).

Normalization, across chips, is how two or more chips are adjusted to compare signal across those chips. There are many different ways to normalize. Pomeroy *et al.* used a simple linear regression model based on an average representative array to scale all other signals in the experiment by the negative inverse regression slope computed as a function of base versus experiment. Irizarry's method uses a Q-spline to fit to the signal's quantiles. The spline-fitting assumes a non-linear relationship and is generally considered more representative of the system's nature (69).

Perfect Match (PM) correction is another parameter that only MAS5 employs. Here, MAS subtracts the ideal mismatch from the Perfect Match (PM). The ideal MM is never greater than the PM for a particular probe pair and therefore, the signal is never negative. RMA does not use MM information, and therefore, does not PM correct.

Summarization is how the signal is computed from .CEL file, which is the file containing coordinate trimmed-mean value of fluorescence for every area of the probe. Affymetrix's first generation algorithm used an average difference of probe pairs, which was later replaced with a 1-step Tukey's biweight method on the log<sub>2</sub> scale. The Li-Wong method used in dChip is based upon fitting a multi-chip model to each probeset. Essentially, Li-Wong estimates the probe response as a function of the overall expression of the array plus some error (70). RMA, however, uses the medianpolish method.

Medianpolish is another multi-chip linear model on the log<sub>2</sub> scale. Current research involving datasets with known outcomes for cross-validation suggest that medianpolish outperforms Tukey's biweight in robustness; meaning the analysis is more stable (i.e. higher entropy) under various perturbation conditions (69).

Within the last two years, additional improvements have been made on Irizarry's RMA algorithm. Specifically, Irizarry and his coworkers have focused their attention on improving the background subtraction step of microarray preprocessing. As part of the microarray preparation, mRNA is labeled and hybridized to the oligonucleotide wafer. The wafer is then scanned by a light-sensing machine which interprets the brightness of a particular spot as a representation of mRNA transcript abundance. A brighter spot is interpreted as a greater abundance of transcript. However, as with all hybridization chemistry, this process is subject to non-specific binding. In particular, oligonucleotides with high guanine and cystine residues in their sequence have higher chances on average of non-specific binding. This is caused by the additional hydrogen bond in purine nucleic acids. Therefore, on average, the overall result of GC rich sequences is a biased overrepresentation of their abundance. Irizarry and his coworkers have incorporated an empirical Bayesian model-based correction for this bias. The result is an improvement to the background correction portion of their RMA algorithm called GC-RMA (71). The GC-RMA algorithm's performance against other leading technologies was recently explored, suggesting that GC-RMA is generally the best overall choice for preprocessing Affymetrix microarrays using probe-level data (72). The study described in this dissertation uses Irizarry's GC-RMA algorithm, which incorporates model-based

background correction, Q-spline normalization, and finally medianpolish summarization for all Affymetrix microarray preprocessing.

In summary, a systems biology approach to solving a biological problem is as much a change in philosophy as it is a technology. Indeed, the philosophy demands certain considerations when one thinks about experimental design: setting up the biological system in such a way as to take advantage of high-throughput measurement technologies, incorporation of sometimes huge datasets measuring many variables at once into a coherent bioinformatic analysis that uses massive amounts of information as an opportunity rather than a hindrance, the necessary informatics and hardware to handle the data, and finally creative use of existing and emerging techniques borrowed from areas of computer science, artificial intelligence and biostatistics.

There are, of course, many high-throughput technologies currently used in systems biology directed research including genomics, proteomics, and metabonomics. By conventional wisdom, everything starts with the gene. Of the three disciplines, genomic technology is the most mature. This is because, among other things, the complete nucleotide sequences of humans and other organisms are known and microarray technologies to measure transcript abundance have been around for several years. Microarray technologies enable measurement of large numbers of genes at the same time. As a consequence, these technologies allow the experimental perturbation of a system in one way or another and observe how the components of the system change. The first problem is how do you define differential gene expression? The second question is once you have defined the set of differentially expressed genes, how do you use that

information to infer novel pathways? And finally, if you are successful inferring novel pathways, you have to have ways to confirm if those relationships actually exist?

This dissertation presents our efforts to answer each of these three questions using many of the above mentioned methods and techniques to address the genomic consequences of the activated nuclear hormone receptor peroxisome proliferator-activated receptor-gamma (PPAR $\gamma$ ) in human colorectal cancer cells. This receptor is a transcription factor, and as such, we hypothesize the phenotype controlled by this receptor is the result of genomic signal transduction. We will present our cell model system and why we feel this system is an excellent model for studying the genomic consequences of PPAR $\gamma$  action. Our cell model system will be first characterized biologically. Cellular phenotypes under the control of PPAR $\gamma$  will be assessed and measured. The system will be perturbed with a synthetic ligand to PPAR $\gamma$  and the resultant signal transduction will be measured by using high-throughput genomics through time. Finally we will explain how these data are used to infer a novel, and previously unappreciated pathway under the control of PPAR $\gamma$ . Finally we will show that manipulation of this pathway leads to certain predicted consequences of PPAR $\gamma$ 's phenotype in the cell model system.

## **PEROXISOME PROLIFERATOR-ACTIVATED RECEPTORS**

### **Nuclear hormone receptors**

The nuclear hormone receptor superfamily is comprised of approximately one hundred transcription factors that are involved in almost every aspect of normal human physiology as well as many human diseases (73). Members of this superfamily bind to specific response elements on target genes as homo- or heterodimers and stimulate or inhibit transcription of these genes by recruiting co-activators or co-repressors. Typical response elements of nuclear hormone receptors are hexanucleotide half-sites separated by variable length of nucleotides between partially conserved direct or inverted half-site repeats. Nuclear hormone receptors can be broadly divided into four categories according to their dimerization and DNA binding properties (reviewed in (73).) Class I receptors include the steroid hormone receptors such as glucocorticoid receptor (GR), progesterone receptor (PR), androgen receptor (AR) and estrogen receptor (ER). These receptors function as homodimers binding to half-site response elements organized as inverted repeats. Class II receptors form heterodimers with retinoid X receptor (RXR) and bind to direct repeats. Members of this class include thyroid hormone receptor (TR), retinoic acid receptor (RAR), Vitamin D receptor (VDR), peroxisome proliferator-activated receptor (PPAR), and liver X receptor (LXR). Most orphan receptors, such as hepatocyte nuclear factor 4 (HNF-4), chicken ovalbumin upstream promoter transcription factor (COUP-TF), steroidogenic factor 1 (SF-1), and retinoid Z receptor (RZR), fall into

class III and class IV. These function as homodimers binding to direct repeats (class III), or monomers binding to single site response elements (class IV).

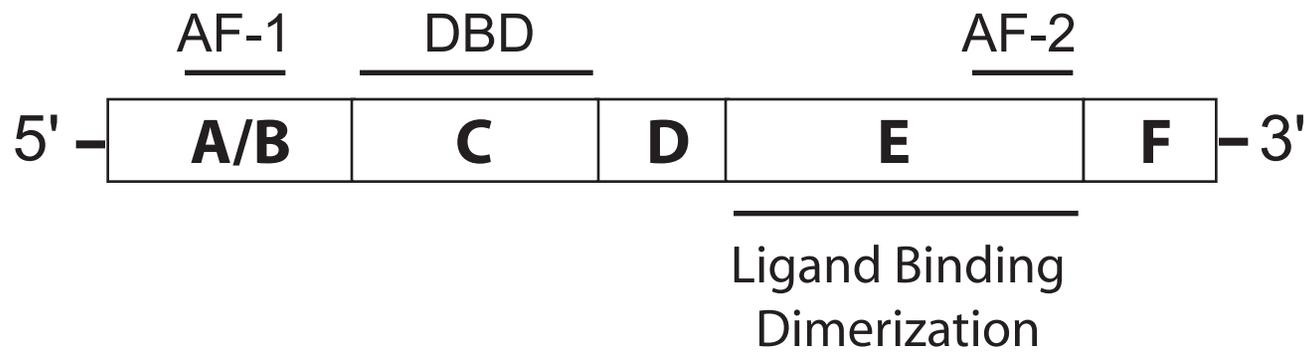
All of the nuclear hormone receptors have common structural features [Figure 3]. They are characterized by a variable N-terminal region (A/B), a highly conserved DNA binding domain (C), a variable hinge region (D), a conserved ligand binding domain (E), and a variable C-terminal region (F) (73). The central DNA binding domain (DBD) is composed of two highly conserved zinc fingers that set the nuclear hormone receptors apart from other DNA binding proteins (74, 75). The DBD is responsible for targeting the receptor to highly specific response elements (73). The ligand binding domain (LBD) is located in the C-terminal half of the receptor. The LBD mediates ligand binding, receptor dimerization and a ligand-dependent transactivation function (activation function 2, AF-2) (76). The A/B region is the most variable both in size and sequence (77). In most cases, it contains an activation function 1 (AF-1) domain, which operates in a ligand-independent, cell and promoter-specific manner (76).

### **PPAR family**

The peroxisome proliferator-activated receptors (PPARs) belong to the nuclear hormone receptor superfamily. The name PPAR came from the initial cloning of one isoform as a target of xenobiotic compounds that induce proliferation of peroxisomes in mouse liver (78). This isoform is now known as PPAR $\alpha$ . Since then, the PPAR family has been expanded to include PPAR $\delta$  (also known as PPAR $\beta$ , NUC1, or FAAR) and

**Figure 3: *Common structural features of nuclear hormone receptors***

A typical nuclear hormone receptor contains seven functional domains. The variable N-region (A/B) contains the ligand-independent AF-1 domain. The DNA binding domain (DBD) (C) is highly conserved and it is responsible for the recognition of specific response elements. The hinge region (D) connects the DBD to the ligand-binding domain (LBD). LBD mediates ligand binding and dimerization. The ligand-dependent AF-2 domain is located at the C-terminal portion of the LBD. [Modified from (79)]



PPAR $\gamma$  (80-86) in various species, including xenopus, mouse, rat, hamster, and human. PPARs form heterodimers with the retinoid X receptor (RXR) and bind to specific DNA sequences known as peroxisome proliferator response elements (PPREs) (87) in the absence of ligand. The conformation of PPAR is altered upon ligand binding, leading to recruitment of transcriptional coactivators/corepressors and an increase/decrease in gene transcription (88).

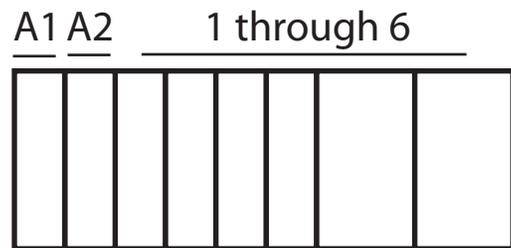
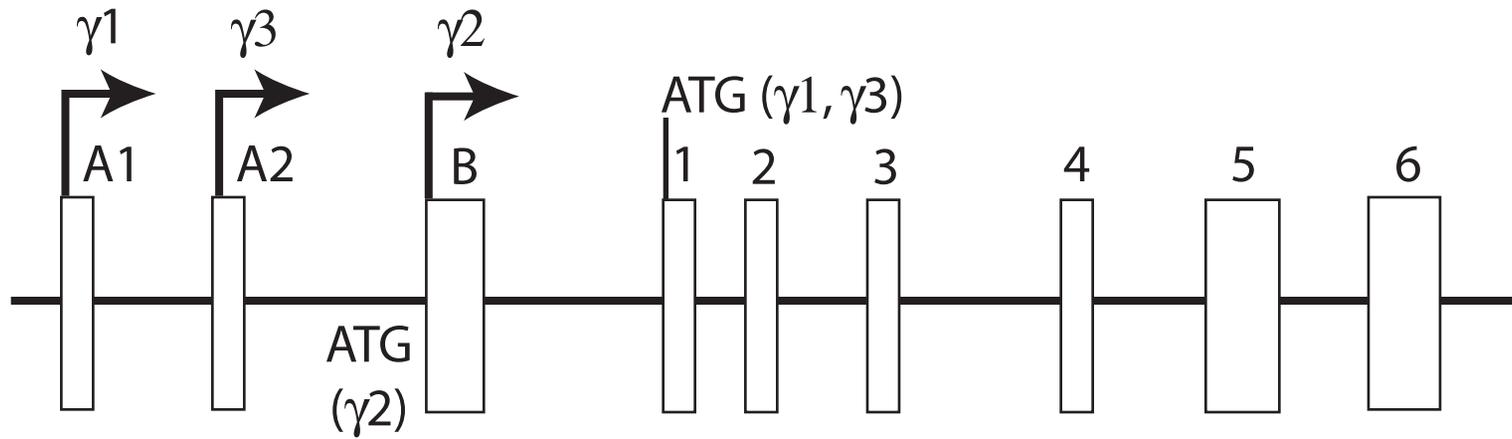
### **PPAR $\gamma$**

In human, the PPAR $\gamma$  gene contains three promoters that yield three RNA isoforms,  $\gamma$ 1,  $\gamma$ 2, and  $\gamma$ 3, via alternative promoter use and different splicing (81, 86, 89-92). Human PPAR $\gamma$ 1 (hPPAR $\gamma$ 1) is encoded by 8 exons, whereas hPPAR $\gamma$ 2 is encoded by seven exons **[Figure 4]**. Exon A1 and A2 comprise the 5'-untranslated region (5'-UTR) of hPPAR $\gamma$ 1. The 5'-UTR and the additional hPPAR $\gamma$ 2 specific 5' sequences are encoded by exon B, which is located between exons A2 and exon 1. The remaining exons, 1 through 6, are common to hPPAR $\gamma$ 1 and hPPAR $\gamma$ 2. Consequently, hPPAR $\gamma$ 2 differs from hPPAR $\gamma$ 1 by having 30 additional N-terminal amino acids. A third subtype, hPPAR $\gamma$ 3 was identified recently (92). The PPAR $\gamma$ 1 and  $\gamma$ 3 RNA transcripts both translate into PPAR $\gamma$ 1 protein.

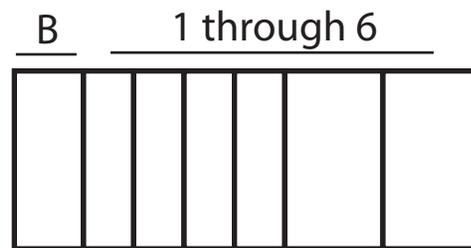
PPAR $\gamma$ 2 is expressed primarily in adipose tissue (90). PPAR $\gamma$ 1 is expressed in a broad range of tissues, including large intestine, small intestine, colon, kidney, muscle,

**Figure 4: *Alternate isoforms of the human PPAR $\gamma$  gene***

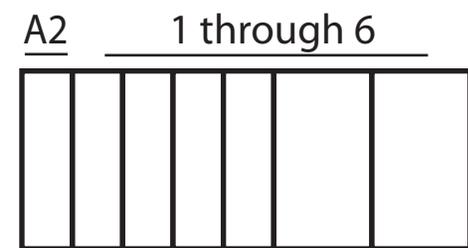
The gene is shown in 5' to 3' orientation and drawn to scale. Exons are denoted by rectangles and introns by a solid line. The location of the ATG start-codon is indicated. Exons 1 to 6 are shared by all three PPAR $\gamma$  isoforms. PPAR $\gamma$ 1 contains in addition the untranslated exons A1 and A2. PPAR $\gamma$ 2 contains exon B, which is translated. PPAR $\gamma$ 3 contains only the untranslated exon A2. [Modified from (79)]



hPPAR $\gamma_1$



hPPAR $\gamma_2$



hPPAR $\gamma_3$

liver, adipose tissue, heart, brain, and pancreas (90, 91). PPAR $\gamma$ 3 expression is restricted to macrophage (93), adipose tissue and colon (92). The different tissue distribution patterns of the three isoforms of PPAR $\gamma$ , together with the presence of multiple promoters, suggest that these different PPAR $\gamma$  isoforms may be important in tissue-specific regulation of gene expression.

Like other members of the nuclear receptor superfamily, PPAR $\gamma$  possess a characteristic modular structure consisting of five functional domains [**Figure 3**] (94). The receptor has a poorly characterized N-terminal region that contains the ligand-independent AF-1 domain. This is followed by a highly conserved DNA binding domain (encoded by exons 2 and 3) that mediates the binding of the receptor to specific response element. At the carboxyl terminus is a dimerization and ligand binding domain, encoded by exons 5 and 6. The LBD contains the ligand-dependent transactivation function AF-2.

The prototypic PPAR $\gamma$  responsive element (PPRE), found in the acyl-CoA oxidase (ACO) gene promoter, contains two copies of the core binding sequence AGG(A/T)CA separated by one nucleotide (87). This conformation, a directly repeated DNA sequence separated by a single nucleotide, forms the so-called DR-1 element. This feature sets PPAR $\gamma$  apart from other receptors that heterodimerize with RXR, but bind to direct repeats separated by multiple nucleotides, such as RAR, VDR, and TR (77). The AGG(A/T)CA consensus is shared by PPAR $\alpha$  (87), but not PPAR $\delta$ . PPAR $\delta$  has its own DR-1 binding consensus CGCTCAC (95).

Naturally occurring ligands of PPAR $\gamma$  include a number of polyunsaturated fatty acids and their derivatives (96). A prostaglandin D-2 derivative, 15-deoxy- $\Delta^{12,14}$ -prostaglandin J<sub>2</sub> (15d-PGJ<sub>2</sub>), was found to be a PPAR $\gamma$  specific ligand (97, 98), activating PPAR $\gamma$  at micromolar concentration. More recently, an oxidized alkyl phospholipid, hexadecyl azelaoyl phosphatidylcholine, was demonstrated to bind PPAR $\gamma$  with a K<sub>d</sub> of ~40 nM and activate the receptor with a similar EC<sub>50</sub> (99). This affinity is the highest reported so far for a natural PPAR $\gamma$  ligand.

A number of synthetic PPAR $\gamma$  ligands have been identified during the past several years, of which the best characterized are the thiazolidinediones (TZDs) (100). TZDs are a class of insulin-sensitizing compounds, which include troglitazone, rosiglitazone, and pioglitazone. The TZDs were initially developed for the treatment of type II diabetes. They are able to enhance insulin actions by increasing insulin sensitivity in target tissues such as liver, muscle, and fat (101). TZDs also efficiently reduce circulating glucose and fatty acids levels in rodent models of diabetes (101). It wasn't until 1995 that TZDs were shown to be PPAR $\gamma$  ligands (102). The order of the PPAR $\gamma$  binding affinity of the TZDs is rosiglitazone > pioglitazone > troglitazone, correlating well with their hypoglycemic activities *in vivo* (103). Notably, emerging evidence suggests that troglitazone acts independently of PPAR $\gamma$  in many functions (104-107). All TZDs appear to have effects that are not mediated by PPAR $\gamma$ , although these effects usually prevail at ligand concentrations in excess of 1  $\mu$ M (104-107).

Our lab has characterized a novel synthetic ligand for PPAR $\gamma$ , RS5444 (108, 109). RS5444 is a TZD. In non-transformed rat intestinal cell line RIE-1, RS5444 activates PPAR $\gamma$  with an EC<sub>50</sub> about 1/20th that of rosiglitazone (109). The very high affinity of RS5444 for PPAR $\gamma$  (EC<sub>50</sub> about 1 nM) minimizes the risk of off target effects, which prevail only at much higher TZD concentrations. RS5444 is orally available and activates PPAR $\gamma$  at dietary delivery rates of less than 1 mg/kg/day (unpublished data from our laboratory), compared to delivery rates of 30-100 mg/kg/day commonly used for rosiglitazone and 100-200 mg/kg/day used for troglitazone. Again, the high affinity of RS5444 minimizes potential off target effects that may complicate interpretation of studies carried out with TZDs *in vivo*.

### **PPAR $\gamma$ and colorectal cancer**

Colorectal cancer is the second leading cause of cancer mortality in Western societies. There are an estimated 106,680 new cases of colon cancer in the United States in 2006, resulting in 55,170 deaths (110). PPAR $\gamma$  is highly expressed in colon tissue (91). The highest levels of receptor expression was found in post-mitotic, differentiated epithelial cells facing the lumen (111). PPAR $\gamma$ 1 is the splice isoform expressed by all the human colorectal cancer cell lines (112). PPAR $\gamma$  ligands have been reported to induce growth arrest and differentiation markers of cultured human colon cancer cells (112). Transplantable tumors derived from human colon cancer cells showed a significant growth reduction when mice are treated with troglitazone (113). Rats treated with the

chemical carcinogen azoxymethane (AOM) develop preneoplastic colonic lesions termed aberrant crypt foci (ACF). Administration of troglitazone significantly reduced the number of ACF lesions (114). Four somatic mutations in the PPAR $\gamma$  gene were described in 55 sporadic colon cancers (115). Each of these mutations was reported to reduce PPAR $\gamma$  function, suggesting that colon cancer be associated with loss-of-function mutations in PPAR $\gamma$ . The development of colorectal cancer is also influenced by prostaglandins (116), which are potential ligands of PPAR $\gamma$ . In mice with mutations in the cyclooxygenase-2 (COX-2) gene or in animals treated with COX inhibitors, decreased production of prostaglandins attenuated or prevented colon cancer development (117, 118).

The effects of PPAR $\gamma$  on colon tumorigenesis *in vivo* remain highly controversial. Activation of PPAR $\gamma$  was shown to promote the development of colon tumors in C57BL/6J-APC<sup>Min/+</sup> mice (119-121). APC<sup>Min</sup> (Min = multiple intestinal neoplasia) is a point mutation in the murine homolog of the adenomatosis polyposis coli gene which increases the risk of multiple intestinal adenomas. In addition, dietary administration of TZDs was shown to increase the number of caecal tumors in wild type C57BL/6 mice (122). On the other hand, a recent study using C57BL/6J-APC<sup>Min/+</sup> mice that were haplo-insufficient for PPAR $\gamma$  (PPAR $\gamma$ <sup>+/-</sup>) and APC<sup>Min/+</sup> mice that carried floxed alleles of PPAR $\gamma$  (PPAR $\gamma$ <sup>FL/+</sup> and PPAR $\gamma$ <sup>FL/FL</sup>) in conjunction with expression of the intestinal-specific Villin-Cre transgene showed that intestinal-specific PPAR $\gamma$  deficiency enhanced tumorigenesis in these mice models (123).

The APC<sup>Min</sup> mice harbor a nonsense mutation in the tumor suppressor gene APC (124). Loss of function mutations in this gene are responsible for the hereditary polyposis syndrome familial adenomatous polyposis and are also thought to be one of the major genetic initiating events for a large percentage of the sporadic colorectal cancers (125).

Heterozygous loss of PPAR $\gamma$  causes an increase in  $\beta$ -catenin levels and a greater incidence of colon cancer when animals are treated with AOM (126). However, mice with pre-existing damage to APC develop tumors in a manner that is insensitive to the status of PPAR $\gamma$  (126). These observations support the hypothesis that PPAR $\gamma$  can suppress colon carcinogenesis but only before damage to the APC/ $\beta$ -catenin pathway. Ras activation, a later event following loss of APC, is critical in mediating transition between early adenoma to intermediate adenoma (127). Ras activation leads to an increase in mitogen-activated protein kinase (MAPK) signaling. PPAR $\gamma$  can be inactivated by phosphorylation at Ser82 by MAPK (128, 129). These observations suggest that PPAR $\gamma$  may function as a tumor suppressor only during the early steps of tumor formation. They also support the role of PPAR $\gamma$  ligands as chemopreventative agents for colon cancer, rather than chemotherapeutic agents.

The differential effects of PPAR $\gamma$  in regulating colon cancer development are likely due to the differences in model systems used. The contradictory results in the literature warrant further investigation of PPAR $\gamma$ 's role in the development of normal and malignant colonic epithelium. An obvious goal is to elucidate the mechanism that

accounts for inhibition of colon cancer cell growth by PPAR $\gamma$  and to exploit this information to develop more efficacious therapeutic strategies to minimize potential side effects such as tumor promotion. As will be discussed in detail, we have undertaken a functional genomic approach to solving the cellular mechanism of action of PPAR $\gamma$  in human early stage colon cancer cells, and our data have inferred a novel regulatory network that involves calcium signaling, the calcium-activated phosphatase calcineurin, and an endogenous calcineurin inhibitor called Down Syndrome critical region 1 (DSCR1).

## **DSCR1, CALCINEURIN AND NFAT**

### **DSCR1 as a tumor suppressor**

DSCR1 was recently identified as a gene encoded by a region of chromosome 21 that appears to contribute to some of the neurological and cardiovascular manifestations of Down Syndrome (130, 131). It is a member of a small family of calsupressin genes (which suppress the activity of the serine-threonine phosphatase calcineurin): DSCR1/MCIP1 on chromosome 21q22.1-22.2, DSCR1L/MCIP2/ZAKI4 on chromosome 6, and DSCR1L2/MCIP3 on chromosome 1. Four alternative DSCR1 promoters have been identified by RACE and cDNA library screenings (132). DSCR1 isoform 1 is transcribed from exon 1 and is highly expressed in brain. Isoform 2 is transcribed from exon 2 and appears to be restricted to fetal tissue. Isoform 3, transcribed from exon 3,

has not been identified in any tissue to date. Isoform 4, transcribed from exon 4, is the principle extra-neuronal form of DSCR1 and is widely expressed in tissues of epithelial origin. Parallel investigations have led to the identification of DSCR1 as a critical regulator of myocyte differentiation, a member of the MCIP family, MCIP1 (133). DSCR1 was also cloned as Adapt78, a gene that conveys resistance to oxidative and  $\text{Ca}^{++}$ -induced stress (134, 135). DSCR1 was also independently cloned as a vascular endothelial growth factor (VEGF)-responsive gene involved in regulation of apoptosis and motility of endothelial cells (136, 137).

Tumor biologists have become interested in DSCR1 as a potential candidate to account for reports that individuals with Down Syndrome are resistant to carcinogenesis (138). Research has recently shown that DSCR1 suppresses tumor growth *in vivo* (139). Initial findings suggest that the tumor-suppressive effects of DSCR1 are mediated through angiogenesis-dependent mechanisms and involve the transcription factor NFATc (137).

### **NFATc as a tumor promoter**

Although Nuclear Factor of Activated T Cells (NFATc) was first identified in activated T cells (140), the role of NFATc has been well defined in endothelial cells and skeletal and cardiac myocytes (141). It is now clear that these transcription factors play very important roles in epithelial cells. NFATc appears to function at composite promoter response elements in concert with AP-1, NF $\kappa$ B, and GATA factors to regulate an array of genes that are involved in proliferation, migration, invasion, and inflammation (141).

Researchers have only recently begun to appreciate the role of NFATc in carcinogenesis. NFATc plays an important role in neovascularization and angiogenesis (141). NFATc also plays a direct role in transformation. A dominant negative form of NFATc1 blocks invasion of breast cancer cells (142, 143), and it has been shown that NFATc is required for integrin-mediated motility and invasion of breast cancer cells (142). NFATc-mediated invasion by breast cancer cells results from induction of COX-2 through an NFATc1/AP1-dependent mechanism (144). NFATc activity is regulated via its calcineurin-dependent phosphorylation state (141).

### **Calcineurin**

Calcineurin/PPP3 is a dimeric serine/threonine phosphatase composed of a catalytic CnA subunit and a regulatory CnB subunit. As the name implies, calcineurin is regulated by  $\text{Ca}^{++}$ , in part by binding of calmodulin and in part by direct binding of  $\text{Ca}^{++}$  to the CnB subunit (145). The major substrate of calcineurin is the NFATc family of transcription factors, which are dephosphorylated by calcineurin in a  $\text{Ca}^{++}$ -dependent manner (141). The phenotype of a CnB mutant mouse is identical to that of the NFATc3/c4 knockout as well as the cyclosporine A treated mouse (141). Such observations suggest that calcineurin is dedicated to regulate the activity of NFATc.

### **The signaling axis of DSCR1-Calcineurin-NFATc**

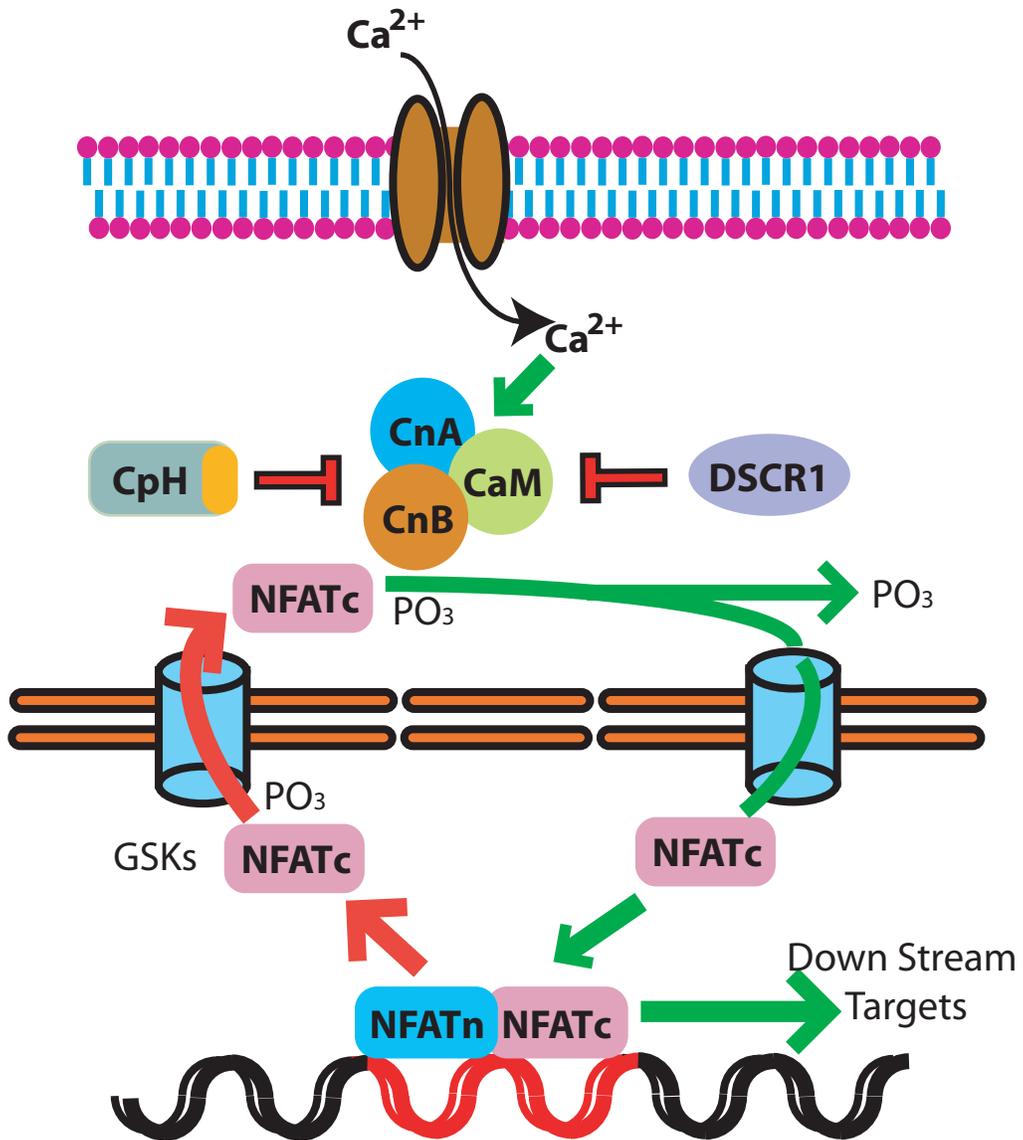
DSCR1 (a.k.a. MCIP1 or Adapt78) is a potent intracellular inhibitor of calcineurin activity (146), and the DSCR1 gene product has sometimes been called calsupressin 1 (147, 148). DSCR1 binds to the active site of calcineurin (133). It was initially reported that inhibition of calcineurin involves sequences in both the N and C termini of DSCR1 (149). The role of N-terminal sequences was counterintuitive, since DSCR1 isoforms have different N-terminal sequences, yet all inhibit calcineurin. A more recent deletion analysis of DSCR1 indicates that inhibition of calcineurin depends entirely upon DSCR1 C-terminal exon 7 (150), which is common to all DSCR1 isoforms (132). A truncated form of DSCR1 isoform 1, which includes exons 1, 5, and 6 does not inhibit calcineurin, whereas exon 7 alone is fully active as a calcineurin inhibitor in vitro and in vivo (150). DSCR1 binds tightly to the active site of calcineurin and prevents binding of the principle calcineurin substrates, the NFATc isoforms (141, 151). The mechanism of calcineurin inhibition by DSCR1 is different from that of the cyclosporine A/cyclophilin A and FK506/FKBP12 complexes, which bind calcineurin at several sites outside of the active site (141). It has been shown that metastatic lesions of thyroid cancer have lost DSCR1 expression, whereas non-metastatic lesions retain DSCR1 (152). This observation suggests that DSCR1 plays a direct role in tumor progression and is consistent with the hypothesis that loss of DSCR1 leads to dysregulation of calcineurin activity and hyperactivation of NFATc in tumor cells.

The calcineurin/NFATc regulatory cycle is illustrated in **[Figure 5]**. Calcineurin-mediated dephosphorylation of NFATc results in unmasking of a nuclear localization

**Figure 5: *Calcium-Calcineurin-NFATc Signaling***

This figure displays the general features of what is known about NFAT signaling. CnA and CnB denote Calcineurin A and B; CaM: calmodulin; CpH and orange oval: cyclosporine bound cyclophilin; DSCR1: Down's Syndrome Critical Region 1. Red denotes inhibition, while green denotes activation. A simplified nuclear pore complex (NPC) is shown; ATP hydrolysis is omitted for simplicity [Modified from (141)]

## Calcium-Calcineurin-NFATc Signaling



signal, which promotes nuclear uptake, and ultimately promoter binding of NFATc. Within the nucleus, NFATc is phosphorylated by a number of serine/threonine protein kinases, resulting in sequestration of the nuclear localization signal and nuclear export (or inhibition of nuclear import) and inhibition of NFATc transcriptional activity (141).

In summary, the signaling axis of DSCR1-calcineurin-NFATc constitutes DSCR1 binding to the active site of calcineurin, which in turn inhibits calcineurin-mediated dephosphorylation of NFATc, blocks nuclear uptake of NFATc, and inhibits transactivation of NFAT target genes, including genes that are involved in proliferation and invasion.

There is limited information about calcineurin/NFATc signaling in gastrointestinal cancer. It has been shown that NFATc1 is required for integrin-mediated transformation of a colon cancer cell line in culture (142). NFATc1/2 induces COX-2 in colon cancer cells (153). COX-2 is a known colon cancer promoter, which is over-expressed in many colon tumors (154) and plays a role in invasion of such cells (155). NFATc1 also promotes pancreatic cancer growth by inducing c-MYC transcription (156). Thus, there is strong evidence that NFATc plays a major role in invasion of colon cancer cell lines, and NFATc may regulate proliferation of such cells, as well. Calcineurin and NFATc1/2 are over-expressed in some primary colon tumors, and such tumors exhibit increased nuclear localization of calcineurin and NFATc, indicative of NFATc activation (153, 157). The data are consistent with the hypothesis that proliferation (e.g. c-MYC) and progression (e.g. COX-2) are stimulated by the calcineurin/NFATc pathway in colon cancer.

One of our goals is to test the hypothesis that induction of DSCR1 inhibits NFATc signaling and accounts for the ability of PPAR $\gamma$  to inhibit proliferation and suppress invasion of colon cancer cells. It is worth emphasizing that PPAR $\gamma$  antagonizes the same cellular processes that are stimulated by NFATc (proliferation, invasion, inflammation), and that the antagonistic effects of PPAR $\gamma$  on these processes have often been linked to crosstalk between PPAR $\gamma$  and AP-1 or NF- $\kappa$ B signaling. The observations that PPAR $\gamma$  and NFATc act in opposition to each other and that this antagonistic effect converges upon a common set of transcriptional regulators imply a mechanistic connection between PPAR $\gamma$  and NFATc signaling. We have used a genomics systems biology approach to test this hypothesis in human colon cancer cells.

## **CHAPTER 2: MATERIALS AND METHODS**

### **CHEMICAL REAGENTS**

The thiazolidinedione RS5444 was provided by Sankyo Ltd., Tokyo, Japan (108). Rosiglitazone (BRL 49653) was purchased from ChemPacific. Fenofibrate and dimethyl sulphoxide (DMSO) was purchased from Sigma. Carbaprostacyclin (cPGI) was purchased from Cayman Chemical. Cyclosporin A was purchased from Sigma. Cyclohexamide was purchased from Calbiochem.

### **PLASMIDS AND VECTORS**

Mouse PPAR $\gamma$ 1 expression vector pSG5-mPPAR $\gamma$ 1 and the PPAR $\gamma$  reporter PPRE<sub>3</sub>-TK-Luc were provided by Dr. Steven A. Kliewer. Murine PPAR $\gamma$ 1 cDNA was amplified through PCR and cloned into the expression vector pSG5 (Stratagene) (158). The PPRE/luc PPAR $\gamma$  reporter was constructed by inserting three copies of an oligonucleotide encoding the PPRE (GTCGACAGGGGACCAGGACAAA-GGTCACGTTTCGGGAGTCAC) in direct orientation into the unique Sall site of the basal reporter construct TK-Luc (159). pCMV-hPPAR $\alpha$ , pCMV-hPPAR $\delta$  and pGL3-NFAT-Luc were gifts from Dr. John A. Copeland (Mayo Clinic, Jacksonville, FL).

## CELL LINES

HEK293 cells were purchased from ATCC maintained in mid log phase growth in DMEM (Invitrogen-Gibco) + 10% FBS (Hyclone). RIE/S3-PPAR $\gamma$ , a non-transformed rat intestinal epithelial cell line stably expressing the murine  $\gamma$ 1 isoform of PPAR, was constructed in our laboratory (109). RIE/S3-PPAR $\gamma$  was maintained in mid log phase growth in DMEM supplemented with 5% charcoal/dextran treated fetal bovine serum (CS-FBS, Hyclone). MOSER S human colorectal cancer cells were obtained from Dr. Raymond Dubois (Vanderbilt University, Nashville, TN). MOSER S cells were maintained in mid log phase growth in DMEM supplemented with 1% FBS. Cells were always maintained below 75% confluence.

For analysis of somatic mutations, genomic DNA was isolated using a QiaAmp DNA minikit (QIAGEN). 1.6 picomolar of primer [Table 1] along with 0.25 $\mu$ g DNA template was brought to a final volume of 6  $\mu$ l and submitted to Mayo Clinic Rochester's Advanced Genomics Technology Center for sequencing using an ABI 3731x1 automated sequencer. Sequencer chromatograph files were analyzed using Chromas Lite software version 2.01 (160).

For the analysis of DSCR1 isoforms, total RNA was isolated using RNAqueous and converted to cDNA (see below). 200 nM of primer [Table 1] along with 160 ng of cDNA was brought to a final volume of 50  $\mu$ l and PCR'd using an Amplitaq Gold kit

Table 1: Primers for MOSER S cells

Name	Sequence	Target
hbeta-cat_exon3_for	GAT TTG ATG GAG TTG GAC ATG G	for sequencing genomic DNA for beta-catenin point mutation exon3
hbeta-cat_exon3_rev	GCT CGA GTC ATT GCA TAC TGT CCA	rev sequencing genomic DNA for beta-catenin point mutation exon3
hK-Ras_exon1_for	GAC TGA ATA TAA ACT TGT GG	for sequencing genomic DNA for K-Ras mutation codon 12 and 13 exon 1
hK-Ras_exon1_rev	TGG TCC TGC ACC AGT AAT ATG C	rev sequencing genomic DNA for K-Ras mutation codon 12 and 13 exon 1
hK-Ras_exon2_for	CTT GGA TAT TCT CGA CAC AGC AGG	for sequencing genomic DNA for K-Ras mutation codon 61 exon 2
hK-Ras_exon2_rev	TAC CGA TGC AGT CTG GAG CAA GTT	rev sequencing genomic DNA for K-Ras mutation codon 61 exon 2
hAPC_exon15A_for	GTG TAG AAG ATA CTC CAA TA	for sequencing genomic DNA for APC codons 1263-1393, exon 15
hAPC_exon15A_rev	GTG AAC TGA CAG AAG TAC AT	rev sequencing genomic DNA for APC codons 1263-1393, exon 15
hAPC_exon15B_for	CAG GGT TCT AGT TTA TCT TC	for sequencing genomic DNA for APC codons 1338-1436, exon 15
hAPC_exon15B_rev	TTC TGC TTG GTG GCA TGG TT	rev sequencing genomic DNA for APC codons 1338-1436, exon 15
hAPC_exon15C_for	GGA ATG GTA AGT GGC ATT AT	for sequencing genomic DNA for APC codons 1412-1515, exon 15
hAPC_exon15C_rev	AAA TGG CTC ATC GAG GCT CA	rev sequencing genomic DNA for APC codons 1412-1515, exon 15
hAPC_exon15D_for	ACT CCA GAT GGA TTT TCT TG	for sequencing genomic DNA for APC codons 1496-1596, exon 15
hAPC_exon15D_rev	GGC TGG CTT TTT TGC TTT AC	rev sequencing genomic DNA for APC codons 1496-1596, exon 15
DSCR1a_for	GTG GCC GGT CCC CAG CT	forward sequencing cDNA for DSCR1 a isoform
DSCR1a_rev	TCT GCA GGT CCA CCT CCT C	reverse sequencing cDNA for DSCR1 a isoform
DSCR1c_for	TTA GCT CCC TGA TTG CCT GT	forward sequencing cDNA for DSCR1 c isoform
DSCR1c_rev	TAC TCC GGC CTC CTG GTC TG	reverse sequencing cDNA for DSCR1 c isoform
DSCR1a,b,c_for	GGG GCC AGG GGA AAA GTA TG	forward sequencing cDNA for DSCR1 a,b,c isoforms
DSCR1a,b,c_rev	TAC TCC GGC CTC CTG GTCT G	reverse sequencing cDNA for DSCR1 a,b,c isoforms

(Applied Biosystems) for 30 cycles using an Applied Biosystems Geneamp 2700 PCR machine. 20  $\mu$ l of PCR reaction was resolved on 1% agarose gel and dyed with 1  $\mu$ g/ml ethidium bromide (EMD chemicals). Gel image was captured on a Kodak 4800 photographic system with an EL Logic 100 camera and Gel Logic 100 imaging software.

## **WESTERN BLOTTING**

Cells were harvested by trypsinization, collected by sedimentation, and lysed in NP-40 lysis buffer (50 mM Tris, 150 mM NaCl, 0.5% NP-40) with protease inhibitor cocktail (Sigma) added just before use. Cell debris was removed by sedimentation and the supernatant reserved. Protein concentration was determined by the method of Bradford (161) using Bio-Rad protein assay reagent (Bio-Rad). Equal amounts (50  $\mu$ g) of protein were resolved by electrophoresis in NOVEX 10% Tris-Glycine gel (Invitrogen) or NuPAGE 3-8% Tris-Acetate gel (Invitrogen) and transferred to a PVDF membrane (Millipore). The membrane was blocked in 5% nonfat milk/TBS containing 0.05% Tween 20 (TBST) overnight at 4°C. Thereafter, the membrane was incubated with antibodies to PPAR $\gamma$  (Santa Cruz Biotechnology), APC (Santa Cruz Biotechnology), or  $\beta$ -actin (Santa Cruz Biotechnology) diluted in 1% nonfat milk/TBST. The membrane was washed three times in TBST for 10 minutes each. The membrane was then incubated with HRP conjugated goat anti-mouse IgG (Santa Cruz Biotechnology), HRP conjugated goat anti-rabbit IgG (Santa Cruz Biotechnology), HRP conjugated bovine

anti-goat IgG (Santa Cruz Biotechnology), or HRP conjugated donkey anti-goat IgG (Santa Cruz Biotechnology) diluted in 1% nonfat milk/TBST for 1 h at room temperature. The membrane was washed three times in TBST for 10 minutes each. Antigen-antibody complexes were detected by the ECL-plus chemiluminescent system (Amersham).

### **TRANSIENT TRANSFECTION ASSAY**

HEK293, RIE/S3-PPAR $\gamma$ , and MOSER S cells were plated at 100,000 cells/well in 6 well plates in DMEM + 10% FBS overnight. Medium was aspirated and replaced with DMEM + 0.1% CS-FBS for 24 h prior to transient transfection protocol. pSG5 and/or pCMV-hPPAR $\alpha$ , pSG5-mPPAR $\gamma$ , and pCMV-hPPAR $\delta$  were co-transfected with SV40-pRL (*Renilla* luciferase transcribed from SV40 promoter, Promega) into 50% to 70% confluent cells using Fugene 6 (Roche) at a DNA:liposome ratio of 1:3. Transfection was carried out in serum free DMEM overnight. Fresh medium was added together with appropriate treatment. Cells were harvested after 24 h. Total cell extracts were prepared for Dual-Luciferase Assay (Promega) according to the manufacturer's instructions using a Veritas Microplate luminometer (Turner Biosystems). The activity of *Renilla* luciferase was used as an internal control. Results are expressed as the mean of triplicate determinations  $\pm$  standard deviation.

## **ANCHORAGE-DEPENDENT AND –INDEPENDENT CELL PROLIFERATION**

MOSER S cells were plated at  $2 \times 10^5$  cells per 60 mm tissue culture plates in DMEM + 5% CS-FBS and allowed to attach over night. Cells were then treated with 0.1% DMSO, 10 nM RS5444, or 1  $\mu$ M rosiglitazone. Three plates of cells from each treatment condition were harvested 1, 3 and 5 days after treatment by trypsinization. The number of cells was counted using a Coulter Counter with the instrument diameter set to 8 microns. Results are expressed as the mean of triplicate determinations  $\pm$  standard deviation.

For anchorage-independent proliferation assay,  $2 \times 10^4$  MOSER S cells were suspended in 0.75% Seaplaque agar (FMC, Philadelphia, PA) in DMEM supplemented with 10% CS-FBS, 1% Penicillin-Streptomycin-Glutamine, 10 mM HEPES, 0.1% gentamycin, and 0.1% DMSO or 10 nM RS5444. Suspended cells were layered over 0.75% agar base in the same medium. After 2 weeks, cells were fixed with 100% methanol. The agarose plates were dried on glass and cells were stained with Geimsa Staining solution (J. T. Baker) for 20 min at room temperature. Stained colonies were washed with running H<sub>2</sub>O and counted by a Kodak 4800 photographic system with an EL Logic 100 camera and Gel Logic 100 imaging software. In brief, 60 mm culture plates with stained cultures were placed on a light box and photographed. Camera zoom was set to 48.5 mm and F-stop to 12.5. Gel Logic 100 software was set to receive white light trans-illumination with an exposure time of 0.068 seconds. These settings brought the entire culture plate into the field of view. Images were digitally captured and saved as 8-

bit color TIFF files. Images were imported into Adobe Photoshop (162) and converted into 1-bit black and white images. The circular field was then cropped to ensure no data outside the field were included in the file. All images were kept at the same size and resolution. The modified files were then imported into MATLAB as matrices (163). MATLAB then summed over the entire matrix, summing the bits representing black dots on the photo. The fraction of colonies represented as black spots on the TIFF image (denoted as 1's) divided by the entire area was scored as the percentage of colony coverage. This technique assumes colony monolayers, i.e. negligible overlap of colonies. Although some overlap of colonies was expected, this error was mitigated by the large number of cells counted.

## **COLONY FORMATION ASSAY**

MOSER S cells treated with or without 10 nM RS5444 for 24 h were plated at  $2 \times 10^5$  cells per 60 mm tissue culture dishes. The cultures were grown in media without RS5444 or DMSO for five days. After media was removed, the cells were washed once with PBS and then fixed with ice cold 100% methanol for 10 min at  $-20^{\circ}\text{C}$ . Fixed cells were washed once with PBS and stained with Geimsa Staining solution for 20 min at room temperature. Stained colonies were washed once with PBS and quantified by the Kodak 4800/MATLAB imaging and quantification method described above.

## **MIGRATION AND INVASION ASSAYS**

MOSER S cells treated with or without 10 nM RS5444 for 24 h were harvested using trypsin, washed once with PBS and suspended in serum free DMEM containing 250 µg/ml heat-inactivated BSA (DMEM/BSA). For migration assays, Transwell chamber (6.5 mm diameter, 8 µm pore size, Costar) membranes were coated with 15 µg/ml collagen I (Vitrogen, Collagen Biomaterials) for 30 min at 37°C.  $1 \times 10^5$  cells suspended in DMEM/BSA were added to the upper chamber and DMEM containing 5% CS-FBS was added to the lower chamber. After incubating for 16 h at 37°C, non-migrating cells were removed from the upper chamber with a cotton swab and cells that had migrated to the lower surface of the membrane were washed once with PBS and stained with 0.2% (wt/vol) crystal violet in 2% ethanol. Migration was quantified by counting cells per square millimeter using bright-field optics. For invasion assays, cells were pre-treated and harvested as described above. Matrigel invasion chambers (8 µm pore size, BD BioCoat) were re-hydrated according to the manufacturer's instruction.  $1 \times 10^5$  cells were added to the chamber and allowed to invade for 6 h. Cells that had invaded were stained as described above. Invasion was quantified by counting cells per chamber. All experiments were done in triplicate.

## **SHRNA GENE SILENCING**

DSCR1 short hairpin RNA (sh-RNA) MISSION™ lentiviral transduction particles were obtained from Sigma Aldrich. The particles sequences were identified by the RNAi Interference Technology Resource (RITR) [Table 2]. Invitrogen Virapower Lentiviral Expression Kit (Sigma) was used to produce lentiviral stock (pLenti vector plus clone) in the packaging cell line 293FT as per the manufacture's protocol. MOSER S cells, no more than 70% confluent and growing in DMEM + 0.1% FBS were infected with virus-containing media at a 1:10 dilution of virus to target cell media, and 0.6µg/ml polybrene. 24 h post infection, MOSER S cells were re-fed with DMEM + 1% FBS. 48 h post-infection, cells were collected for accessing knockdown efficiency of clones by quantitative real time PCR (qPCR).

## **RNA ISOLATION**

Total RNA was extracted using RNAqueous (Ambion), according to the manufacturer's protocols. RNA quality and integrity was measured using an Agilent 2100 Bioanalyzer (Agilent Technologies) along with the Agilent RNA 6000 Nano Kit (Agilent Technologies) following the manufacture's protocols. Heat denatured Ambion RNA 6000 sizing ladder was used as the reference. Heat denatured RNA samples (diluted to a concentration of 25 to 250 ng/µl) were loaded onto the Nano chip. The electrophoretic assay was run using Agilent 2100 Expert software accompanying the

Table 2: DSCR1 shRNA MISSION™ lentiviral clone candidates

Gene	TRC Number	Candidate Sequence
Symbol: DSCR1 RefSeq: NM_004414 Organism: Human Taxon ID: 9606	TRCN0000019844	CCGGGCTTCAAACGAGTCAGAATAACTCGAGTTATTCTGACTCGTTTGAAGCTTTTT Clone ID: NM_004414.5-386s1c1 Accession Number(s): NM_004414.5, NM_203418.1, NM_203417.1 Region: CDS
	TRCN0000019845	CCGGCCTCTTTAGGACGTATGACAACCTCGAGTTGTCATACGTCCTAAAGAGGTTTT Clone ID: NM_004414.5-339s1c1 Accession Number(s): NM_004414.5, NM_203418.1, NM_203417.1 Region: CDS
	TRCN0000019846	CCGGCCAAATCCAGACAAGCAGTTTCTCGAGAACTGCTTGTCTGGATTTGGTTTT Clone ID: NM_004414.5-532s1c1 Accession Number(s): NM_004414.5, NM_203418.1, NM_203417.1 Region: CDS
	TRCN0000019847	CCGGGTCCATGTATGTGAGAGTGATCTCGAGATCACTCTCACATACATGGACTTTTT Clone ID: NM_004414.5-709s1c1 Accession Number(s): NM_004414.5, NM_203418.1, NM_203417.1 Region: CDS
	TRCN0000019848	CCGGGAGCTTCATTGACTGCGAGATCTCGAGATCTCGCAGTCAATGAAGCTTTTT Clone ID: NM_004414.5-219s1c1 Accession Number(s): NM_004414.5 Region: CDS

equipment. Both the virtual gel-like banding images and the electropherograms were visually inspected for RNA degradation. Likewise, the 18S/28S ribosomal RNA ratio and a proprietary RNA integrity number (RIN), which incorporates not only the relative ribosomal band strengths, but also elements of the entire electrophoretic trace, were used as an objective measurement. The RIN spans from 1 (complete degradation) to 10 (perfectly intact RNA). Any sample with a RIN number less than 8.5 was discarded from further analysis.

## **QUANTITATIVE REAL-TIME PCR (QPCR)**

The abundance of individual mRNAs was determined using two-step quantitative reverse transcriptase-mediated qPCR. Reverse transcription of total RNA was performed using the High-Capacity cDNA Archive kit (Applied Biosystems). qPCR reactions were performed using 10 ng of input cDNA for both target genes and endogenous controls using the TaqMan Universal PCR master mix (Applied Biosystems). Amplification data were collected using an Applied Biosystems Prism 7900 sequence detector and analyzed using the Sequence Detection System software (Applied Biosystems). Gene targets and TaqMan Assay-On-Demand assays were purchased from Applied Biosystems [**Table 3**]. AB Prism 7900 cycling parameters were set to: AmpErase UNG Activation, 50°C for 2 min; AmpliTaq Gold activation, 95°C for 10 min; denaturation, 95°C for 15 s; and annealing/extension, 60°C for 1 min for 40 cycles. Data were normalized to GAPDH and mRNA abundance was calculated using the  $\Delta\Delta C_T$  method (164).

Table 3: qPCR Applied Biosystems TaqMan® Assays

<b>Gene</b>	<b>AB Assay</b>	<b>RefSeq</b>
hDSCR1	Hs00231766_m1	NM_203417.1, NM_203418.1, NM_004414.5
hF2R	Hs00169258_m1	NM_001992.2
hGAPDH	Hs99999905_m1	NM_002046.2
rGAPDH	Rn99999916_s1	NM_017008.2
hKLF4	Hs00358836_m1	NM_004235.3
hPDK4	Hs00176875_m1	NM_002612.2
rPDK4	Rn01646918_g1	NM_053551.1
hPPARG	Hs00234592_m1	NM_138711.2, NM_138712.2, NM_005037.4, NM_015869.3
hPTGS2	Hs00153133_m1	NM_000963.1
hREG1A	Hs00602710_g1	NM_002909.3
hRHOB	Hs00269660_s1	NM_004040.2
hRUNX2	Hs00231692_m1	NM_001015051.1, NM_001024630.1, NM_004348.3
hTSC22D1	Hs00234686_m1	NM_006022.2
hVEGF	Hs00173626_m1	NM_001025366.1, NM_001025367.1, NM_001025368.1, NM_001025369.1, NM_001025370.1, NM_003376.4

h: human, r: rat

## ESTIMATION OF RS5444'S EC50

Estimation of a drug's EC50 via the measurements of a response utilizes Clark's theory of occupancy. This theory makes the assumption that there exists a linear relationship between an agonist's occupation of receptor binding sites and the downstream response (165).

This idea can be represented as:



In this equation, L represents the amount of ligand, R represents the amount of receptor, LR represents the amount of bound ligand to receptor, and  $k_e$  represents the binding constant. Therefore, if one assumes that the binding constant  $k_e$  is known, then for any concentration [LR], the response is known. From the law of mass action, and receptor occupation [LR] can be expressed as follows:

$$[LR] = \frac{[L][R]_{\text{total}}}{K_D + [L]} = \frac{[L][LR]_{\text{max}}}{K_D + [L]}$$

Clark's theory of occupancy predicts that the maximal response  $\Delta_{\text{max}}$  will occur at  $[LR]_{\text{max}}$  and that at any fraction of  $[LR]_{\text{max}}$ , the response will be proportional to the maximal response. Therefore,

$$\frac{\Delta}{\Delta_{\text{max}}} = \frac{[LR]}{[LR]_{\text{max}}} = \frac{[L]}{K_D + [L]}$$

From this relationship, it also can be shown that the half-maximal response will occur when  $[L] = K_D$ . That is, the concentration of ligand will equal the disassociation

constant. Graphically, this is illustrated by finding the intersection where the ligand concentration equals half of the maximal response. This concentration, by Clark's theory of occupancy, is the effective concentration at 50% of ligand (EC50).

## **MICROARRAY ANALYSIS**

Gene profiling analysis was performed using HGU-133plus2 GeneChip Microarrays (Affymetrix). First-strand cDNA synthesis was performed using total RNA (10 – 25 µg), a T1-(dT)24 oligomer (5'GGCCAGTGAATTGTAATACGACTCACTAT GGGAGGCGG-dT24 3') and SuperScript II reverse transcriptase (Invitrogen). The cDNA was converted to double-stranded DNA by transcription *in vitro*. cRNAs were synthesized using bacteriophage T7 RNA polymerase in the presence of biotinylated nucleotides. Biotin-labeled target RNAs were fragmented to a mean size of 200 bases. Hybridization was performed at 45°C for 6 h in 0.1 M morpholinoethane sulfonic acid (MES), pH 6.6, 1 M sodium chloride (NaCl), 0.02 M ethylenediaminetetraacetic acid (EDTA), and 0.01% Tween 20. Arrays were washed using both nonstringent (1 M NaCl, 25°C) and stringent (1 M NaCl, 50°C) conditions prior to staining with phycoerythrin labeled streptavidin (10 µg/ml). Data were collected using a Gene Array Scanner (Hewlett Packard) and raw .CEL files were extracted using Affymetrix GeneChip® Operating Software (GCOS). All experiments were done in triplicate.

## **QUALITY CONTROL AND PREPROCESSING OF AFFYMETRIX**

### **MICROARRAY PROBE-LEVEL DATA**

Affymetrix .CEL files, extracted from Affymetrix GCOS software, were loaded into the R language environment (version 2.2.1) (166) with the Bioconductor (167) libraries (version 1.7) compiled from source and run under the Linux distribution OpenSuse v.10 64-bit operating system with an AMD 3.5 GHz CPU, 4 gigabytes of RAM and 5 gigabytes of swap space. CEL image files were displayed for visual inspection. If any image displaying chip artifacts such as scratches, ghosting, or uneven liquid pooling were found, the replicate would be excluded from further analysis. Next, computational preprocessing was performed on the samples. Two separate groups were analyzed as complete datasets. The first dataset included the entire set of genechips in the presence of RS5444 over the course of 24 h in triplicate every two hours (0 h, 2 h, 4 h,... 24 h) totaling 39 HGU-133plus2 gene chips. The second dataset included the entire set of genechips in the presence of RS5444 or 0.1% DMSO in triplicate every six hours (0 h +/- RS5444, 6 h +/- RS5444, ... 24 h +/- RS5444) totaling 30 HGU-133plus2 gene chips. Both datasets were preprocessed using several algorithms to obtain different figures of merit. The first metric calculated was Bioconductor's implementation of Affymetrix MAS 5.0 change call using the `mas5calls` function with default parameters under the `affy` package (version 1.8.1) (168). The second metric was Bioconductor's implementation of Affymetrix MAS 5.0 summarization values using the `mas5` function under the `affy` package with default parameters (168). The default parameters of the

mas5 function instruct the function to neither normalize nor background correct, exactly mimicking data treatment under the Affymetrix MAS 5.0 software suite. The third metric was Irizarry's robust multiarray average (RMA) summarization values (log base 2 output) using the justRMA function call with default parameters under the affy package (69). The default parameters instruct the algorithm to both background adjust the data and normalize by quantiles. Finally, the fourth metric was Wu and Irizarry's GC-RMA summarization values (log base 2 output) using the justGCRMA function call with default parameters under the package GCRMA (version 2.2.0) (71). The default parameters under justGCRMA instruct the algorithm to both background adjust using elements of RMA functionality augmented with a full affinity model of GC content of each oligonucleotide and then normalize by quantiles. The datasets were then exported into Insightful Splu 7 with ArrayAnalyzer 2.0. Both Bland-Altman (MvA) plots and principle components analysis (unsupervised clustering) were calculated to determine more refined analysis of across-replicate and across-time variation of the gene chips (169). Summarization data were then imported into a Microsoft 2003 Access database. Affymetrix probesets were then filtered based on change call information. That is, only probe sets that were determined by the Bioconductor mas5calls function to be either present "P" or marginally present "M" across all conditions were used for further analysis. Any Affymetrix probe set that contained at least one absent call "A" was excluded from further analysis.

## **STATISTICAL MODELS FOR DIFFERENTIAL EXPRESSION OF LONGITUDINAL MICROARRAY DATA**

Three different statistical models were considered for scoring significant differential expression, each making different assumptions about the data. The first was a non-parametric model making no assumptions at all about the input data structure or possible underlying biological meaning. The second model was slightly more sophisticated. Here, distances were computed based on a two step least-squares regression fit to find probe sets with significant temporal expression changes and significant differences between experimental groups. The third model employs the most sophisticated set of assumptions. The generalized multivariate t-test, known as the HotellingT2 test, was modified by inferring from the data hyper-parameters (tuning parameters). These tuning parameters adjust the co-variance matrix and ultimately reduces both type-I and type-II statistical error.

### **Non-parametric model**

A non-parametric model was implemented in MATLAB version 6.5.1 (163) which takes the two-class longitudinal dataset as input and returns as output a list of p-values for each probe-set. Significance was calculated as a test of the alternate hypothesis (i.e. the median or maximum distance from a particular probe set's control kinetic profile versus the treatment kinetic profile) versus the same probe set's null distribution (null hypothesis). A script then calls the function `mt.rawp2adjp` from the

package `multtest` in the R language environment (170). This function returns a matrix of several different adjusted p-values (e.g. Bonferroni, Benjamini & Hochberg, etc.) for each of the probe set's raw p-values. As the model's title implies, the model makes no assumptions *a priori* of the data's underlying biological meaning. The algorithm expects as input for each Affymetrix probe set triplicate measurements in the presence and absence of drug over five time points (0 h, 6 h, 12 h, 18 h, and 24 h).

*Pseudo-Code of non-parametric model:*

C = control; D= drug

1. Calculate distance measurement for alternate hypothesis. For any two pairs of data C & D such that:

- a.  $C_i : C_0, C_6, C_{12}, C_{18}, C_{24}$  in triplicate  $C_j : \{X_1, X_2, X_3\}$
- b.  $D_i : D_0, D_6, D_{12}, D_{18}, D_{24}$  in triplicate  $D_j : \{D_1, D_2, D_3\}$
- c.  $D_i \leftarrow \{D_i, C_i\}$
- d.  $C_i \leftarrow \{\text{compliment}\}$
- e. For all i,  $|\text{Dist}_{\max}| = \text{Max}_j \{|D_j - C_j|\}$

2. Calculate distance measurement for null hypothesis. For  $i = \{0, 6, 12, 18, 24\}$

- a.  $D_i^* \leftarrow \{D_i, C_i\}$
- b.  $C_i^* \leftarrow \{\text{compliment}\}$

c. For all  $i$ ,  $|\text{Dist}_{\max}^*| = \text{Max}_j \{ |D_i^* - C_i^*| \}$

d.  $|\text{Dist}_{\max}^*| \Rightarrow$  null histogram

3. Repeat step (2) 1000 times

4. Calculate raw p-value:  $P_{\text{value}} = \# \frac{\{ |\text{Dist}_{\max}^*| \leq |\text{Dist}_{\max}| \}}{\beta}$

5. For each p-value, call R library through DCOM server for multiple test adjustment

a.  $\text{fn}(\text{mt.rawp2adjp}(\text{raw p-value})) =$  adjusted p-value

### Polynomial Regression Model

The model is part of the bioconductor library *maSigPro* and utilizes a two-step regression approach to score significant temporal expression changes and significant changes between experimental groups. The method defines a general regression model for the data where the experimental groups are identified by dummy variables. The procedure first adjusts this global model by the least-squared technique to identify differentially expressed genes and selects significant genes applying false discovery rate control procedures. Secondly, stepwise regression is applied as a variable selection strategy to study differences between experimental groups and to find statistically significant different profiles. The coefficients obtained in this second regression model will be useful to cluster significant genes with similar expression patterns and to visualize the results.

A full explanation of this methodology is covered by Conesa *et al.* (171). Briefly, two kinds of variables will be considered: expression which will be an independent and continuous and time which will be dependent and discrete. Therefore, let there exist  $I$  experimental groups describe by the qualitative variable at  $J$  time points for each particular condition  $j$ , that is ( $i = 1, \dots, I$  and  $j = 1, \dots, J$ ). Gene expression (probe set expression) will be measured for  $N$  genes over  $R_{ij}$  replicates. Likewise there will be  $I - 1$  dummy variables to distinguish between each experimental condition.

Let  $y_{ijr}$  denote the normalized and transformed expression value from each gene under the condition  $ijr$ , where replicate  $r = (1, \dots, R_{ij})$ . Therefore, to explain the evolution of  $y$  along the time ( $T$ ) we consider the following polynomial model, where simple time effects and interactions between the dummy variables and time have been modeled. In principle, the maSigPro methodology allows a polynomial model of  $J - 1$  degrees:

$$\begin{aligned}
y_{ijr} = & \beta_0 + \beta_1 D_{1ijr} + \dots + \beta_{(1-1)} D_{(1-1)ijr} \\
& + \delta_0 T_{ijr} + \delta_1 T_{ijr} D_{1ijr} + \dots + \delta_{(1-1)} T_{ijr} D_{(1-1)ijr} \\
& + \gamma_0 T_{ijr}^2 + \gamma_1 T_{ijr}^2 D_{1ijr} + \dots + \gamma_{(1-1)} T_{ijr}^2 D_{(1-1)ijr} \\
& \dots \\
& + \lambda_0 T_{ijr}^{J-1} + \lambda_1 T_{ijr}^{J-1} D_{1ijr} + \dots + \lambda_{(1-1)} T_{ijr}^{J-1} D_{(1-1)ijr} + \varepsilon_{ijr}
\end{aligned}$$

where  $\beta_0, \delta_0, \gamma_0, \dots, \lambda_0$  are the regression coefficients representing the reference group.

$\beta_i, \delta_i, \gamma_i, \dots, \lambda_i$  are the regression coefficients that account for specific polynomial differences (linear, quadratic, cubic, etc.) between the  $(i + 1)$ -th group profile and the first

group (reference) profile,  $i = 1, \dots, I - 1$ . The term  $\varepsilon_{ijr}$  is the random variation associated with each gene at measurement  $ijr$ .

With this general polynomial model of each gene under each condition, the algorithm first defines the null and alternate hypothesis, calculating an ANOVA for each gene. Statistically significant genes will show a correspondingly significant F-statistic, which is then corrected for multiple testing using Benjamini-Hochberg False Detection Rate correction (172). The second step is to take significant genes as defined by their adjusted p-value, the regression coefficients from these genes are then used to determine the conditions in which those genes show statistically significant profile changes utilizing a variable selection strategy (stepwise regression discussed by Draper and Smith (173)).

### **Empirical Bayesian Model**

This model is part of the bioconductor library *timecourse* and considers replicate preprocessed summarization data for one gene  $X$ , where there are  $n$  replicate measurements  $X_1, \dots, X_n$ . Therefore, the null hypothesis would be stated as  $H_0 : \mu_X = 0$  and the alternate hypothesis is  $H_1 : \mu_X \neq 0$ . One way to test this difference could be the application of the Student's t-test, where the t-statistic is defined as:

$$t = \bar{X} - \frac{\text{null value}}{\text{SE}(\bar{X})} = \frac{\bar{X} - \sigma}{\text{SE}(\bar{X})}$$

$$\text{SE}(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

where.

$SE(\bar{X})$  is the standard error of  $(\bar{X})$

$\bar{X}$  is the average of  $X_1, \dots, X_n$

$n$  is the number of replicates for each  $X$

$\sigma$  is the variance of the population

This simple statistic is for the specific case in which one wishes to test an observed dataset containing one subject (one gene in this case) over  $n$  replicates. This framework then allows a method to test differences between two populations. In the case of genes, for example, one then has a framework to define differences between a gene in one condition (control) versus a different condition (treatment) and calculate a metric for this difference, i.e. the t-statistic. In this case, the null hypothesis would be one population, with its mean centered around zero, and the alternate hypothesis' premise is that the second population's mean is centered somewhere other than zero.

Consider now, the more general case where the dataset contains  $m$  subjects (many genes), each subject having  $n$  replicates. The notation for this is represented as a  $m$ -by- $n$  matrix:

$$\begin{pmatrix} X_{11} & \dots & X_{1n} \\ \vdots & \ddots & \vdots \\ X_{m1} & \dots & X_{mn} \end{pmatrix}$$

Each gene  $X_1, \dots, X_m$  has  $n$  replicates. In the case of multiple subjects (genes) in matrix form, the replicates are noted as vectors. For example, the first gene's value over 1 to  $n$  replicates in vector notation is  $\underline{X}_1 = X_{11}, \dots, X_{1n}$  and the average value of the replicates for the same gene is  $\bar{X}_1$  in vector notation. Therefore, in matrix notation, the generalized

notation of averages for all genes is  $\bar{\underline{X}} = \bar{X}_1, \dots, \bar{X}_m$ . Likewise, just as in the univariate case, the null hypothesis is  $H_0 = (\mu_1, \dots, \mu_m) = (0_1, \dots, 0_m)$  and the alternative hypothesis is  $H_0 \neq (\mu_1, \dots, \mu_m) \neq (0_1, \dots, 0_m)$ . As a comparison for the univariate and multivariate case observe that:

$$\begin{aligned} \text{Univariate:} \quad t &= \frac{\bar{X} - \text{null value}}{\text{SE}(\bar{X})} \\ \text{Multivariate:} \quad \underline{t} &= \frac{(\bar{X}_1, \dots, \bar{X}_m) - (0_1, \dots, 0_m)}{\text{SE}(\bar{\underline{X}})} \end{aligned}$$

Again, in the generalized case of multivariate data, the vector notation for standard error is  $\text{SE}(\bar{X}) = \sqrt{\text{variance}(\bar{X})}$ . Likewise, the variance for a matrix is denoted as covariance, where each value has a variance value against itself and every other value. Therefore, in matrix notation, the covariance of matrix  $y$  is shown as:

$$\text{cov}(\underline{y}) = \begin{pmatrix} \sigma^2_1 & \cdots & \gamma_{1m} \\ \vdots & \ddots & \vdots \\ \gamma_{1m} & \cdots & \sigma^2_m \end{pmatrix}$$

Again, following from the univariate case, the t-statistic for the generalized case of a  $m$ -by- $n$  matrix would be:

$$\underline{t} = S^{-1/2} (\bar{X}_1, \dots, \bar{X}_m) - (0_1, \dots, 0_m) n^{-1/2}$$

where.

$$S = \text{cov}(\bar{X}_1, \dots, \bar{X}_m)$$

$$\text{variance}(\bar{X}) = \left( \frac{\sigma^2}{n} \right)$$

$$\text{SE}(\bar{X}) = \left( \frac{\sigma}{\sqrt{n}} \right)$$

For practical reasons, most square the generalized t-statistic leading to the better known HotellingT2 metric, where T2 signifies the squaring:

$$\underline{t}^2 = n^{1/2} \tilde{S}^{-1/2} (\bar{X} - \mathbf{0})$$

$$\underline{t}^2 = n^{1/2} \tilde{S}^{-1/2} (\bar{d})$$

It is from this point where Tai and Speed introduce an inferred tuning *hyperparameter* to the HotellingT2 statistic which attempts to *moderate* the model. The hyperparameter matrix is estimated from the entire data set using an empirical Bayes learning algorithm. Moderation in this context refers to a correction for both false negatives and false positives arising from poorly estimated covariance relationships. Normally, large changes in gene expression over time would be logical targets to further study. However, given the enormous amount of variables to be estimated, relative to the number of time points available for estimation large errors in replicate variance-covariance relationships are virtually guaranteed. Adding to the problem, small changes in expression over time with small replicates can lead to large between and within F-statistics; a condition caused by dividing the statistic by very small values. As a result, Tai and Speed propose a new inferred correlation matrix  $\underline{S}$  that corrects for this model

artifact by *moderating* the statistic (protecting from wild fluctuations) through a learned model (174).

$$\tilde{\mathbf{S}} = \frac{(n-1)\mathbf{S} + \nu\mathbf{\Lambda}}{(n-1) + \nu}$$

where.

$\mathbf{S}$  is the gene-specific variance-covariance matrix

$\nu$  and  $\mathbf{\Lambda}$  are hyperparameters estimated via empirical Bayes

Moreover, the Bayesian context of the learned model provides a monotonically increasing posterior odds metric (referred to as the MB statistic) which functions much like a p-value in univariate analysis for determining cut-off thresholds for significantly scored targets (175).

$$\lambda = \frac{P(H_0 | \text{observed data})}{P(H_1 | \text{observed data})}$$

where.

$\lambda$  is the Bayes Factor or Posterior Odds Ratio

$P(x | y)$  is the posterior probability of x conditioned on y

$H_0$  is the null hypothesis

$H_1$  is the alternate hypothesis

## GENE ONTOLOGY ANALYSIS

It is often the case that the biological context in which genes are regulated is more informative than the genes themselves. The gene ontology analysis program ermineJ version 2.1.12 was used to calculate the over-representation of gene ontology terms (176,

177). Longitudinal microarray data were used to define groups of “GO terms” which consist of functionally related genes. Over-representation analysis uses a cut-off threshold to determine if there are gene sets within GO classes which are statistically overrepresented. Over-representation is defined as statistical significance of a particular gene-ontology class by calculating its relative over-representation compared to other classes within a predefined set of genes. We used over-representation analysis instead of the more generalized re-sampling calculation of the functional class score (FCS) (26) for statistical confidence of functional groupings. For this analysis 6127 GO classes were tested for biological function only. This was done because we were extremely confident about our threshold level (see below). Only classes with a minimum of 5 and maximum of 100 genes were taken into account, as very small or very large classes are unlikely to be as informative (either too specific or too general) (178). The data input file consisted of the initial 10227 probe sets with a MAS5 change call of either “P” or “M” across all conditions. In addition, associated HotellingT2 scores were used as the figure of merit to calculate p-values for GO classes. A HotellingT2 score cut-off of 281.7 (corresponding to an MB-statistic cut-off of  $< -2.0$ ) was used as the figure of merit in ermineJ to calculate p-values. Only GO classes with a p-value of 0.05 or less was considered for further analysis.

## CHAPTER 3: RESULTS

### BIOACTIVITY OF RS5444, A HIGH AFFINITY PPAR $\gamma$ SPECIFIC AGONIST

We carried out an initial set of experiments to characterize the novel, third generation thiazolidinedione RS5444. Our initial characterization was performed with RIE/S3-PPAR $\gamma$ , a transgenic cell line derived from RIE/S3 (normal rat intestinal epithelial cells) which stably expresses mouse PPAR $\gamma$ 1 (109). These data demonstrate that RS5444 activates PPAR $\gamma$  using a transiently transfected PPRE-driven luciferase reporter. We next determined whether RS5444 was a PPAR $\gamma$  specific agonist. HEK293 cells, which do not contain functional PPAR isoforms but do contain functional RXR, were transiently transfected with an expression vector for PPAR $\gamma$ , PPAR $\alpha$ , or PPAR $\delta$ , RS5444 only activated PPAR $\gamma$ , but not PPAR $\alpha$  or PPAR $\delta$  [**Figure 6A**]. Under the same experimental conditions PPAR $\alpha$  and PPAR $\delta$  were activated by their own specific agonists (fenofibrate for PPAR $\alpha$  and carbaprostacyclin a.k.a. cPGI for PPAR $\delta$ ).

The EC<sub>50</sub> of RS5444 for the PPAR $\gamma$  receptor was estimated by measuring the induction of an endogenous PPAR $\gamma$  target gene, pyruvate dehydrogenase kinase 4 (PDK4) (109, 179). RIE/S3-PPAR $\gamma$  cells and MOSER S cells (a human colon cancer cell line, discussed in detail in the next section) were treated for 24 h with 0.1% DMSO, 0.1 nM, 1 nM, 5 nM, 10 nM, 50 nM, 100 nM RS5444 or 1 nM, 10 nM, 50 nM, 100 nM, 500

nM, 1  $\mu$ M rosiglitazone. qPCR measuring PDK4 mRNA abundance following the treatment showed that both PPAR $\gamma$  agonists induced PDK4 expression in RIE/S3-PPAR $\gamma$  cells and MOSER S cells to a similar degree. However, RS5444 activated PDK4 transcription at much lower concentration than rosiglitazone [**Figure 6B**]. The EC<sub>50</sub> of RS5444 and rosiglitazone for PPAR $\gamma$  receptor in RIE/S3-PPAR $\gamma$  cells and MOSER S cells were estimated using a model described in the Material and Methods. PDK4 mRNA transcript abundance values were used as input for a single-binding affinity model assuming saturation (76) using SigmaPlot version 9.01 [**Figure 6C**]. The model interpolates 250 values between the input values [**Table 4**] and scores how well the observed response fits to the interpolated response [**Table 5**]. Both the R-squared and residual sum of squares metrics showed good agreement between the observed response and the estimated response, suggesting that the observed data fit the assumption of the single-binding affinity model. Under these conditions, the EC<sub>50</sub> of RS5444 and rosiglitazone is equivalent to their estimated K<sub>d</sub> calculated from Clark's theory of occupancy. Therefore, the observed EC<sub>50</sub> of RS5444 induction of PDK4 transcription in RIE/S3-PPAR $\gamma$ 1 cells and MOSER S cells were both near 0.6 nM [**Table 5**]. Rosiglitazone exhibited an EC<sub>50</sub> of 11.3 nM in RIE/S3-PPAR $\gamma$ 1 cells and 14.5 nM in MOSER S cells [**Table 5**]. These data indicate that RS5444 is a high potency PPAR $\gamma$  specific agonist. This drug was subsequently used at 10 nM in the studies of MOSER S cellular physiology and genomic profiling upon PPAR $\gamma$  activation.

**Figure 6: Bioactivity of a high affinity PPAR $\gamma$  agonist RS5444**

**A)** HEK293 cells were transiently co-transfected with PPRE3-TK-Luc, phRL-SV40, and pSG5 empty vector, or vector expressing human PPAR $\gamma$ , PPAR $\alpha$ , or PPAR $\delta$ . Cells were treated with 0.1% DMSO, 10 nM RS5444, 50  $\mu$ M fenofibrate, or 20  $\mu$ M cPGI for 24 h. Normalized PPRE3-TK-Luc activity was determined and plotted as fold activation relative to cells treated with DMSO. Data represent the mean  $\pm$  SD, n = 3. **B)** RIE/S3-PPAR $\gamma$  cells and MOSER S cells were treated with various concentrations of RS5444 (square) or rosiglitazone (circle) for 24 h. The abundance of PDK4 mRNA was measured by quantitative RT-PCR (qPCR) and normalized to human GAPDH. Data represent the mean  $\pm$  SD, n = 3. **C)** Values denoting fold change of PDK4 transcript abundance relative to GAPDH were used as input for single binding affinity model with saturation to estimate the EC50s of RS5444 and rosiglitazone by Clark's assumption. Data represents modeled (projected) values of PDK4 transcript abundance.

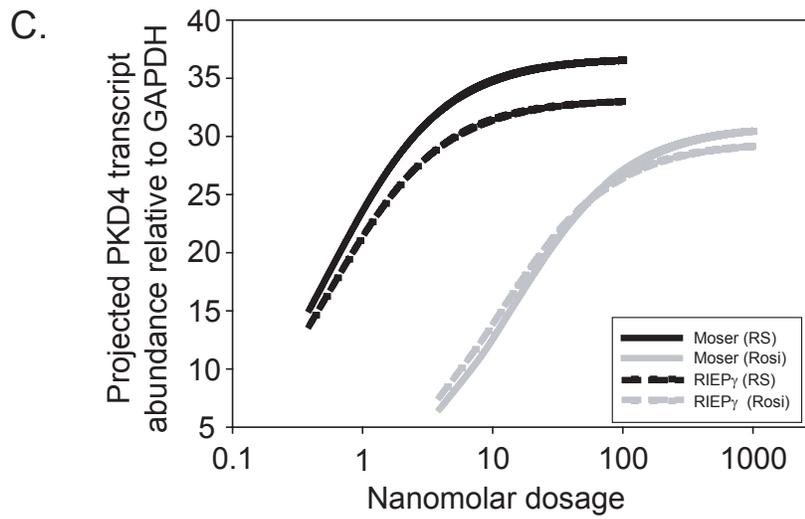
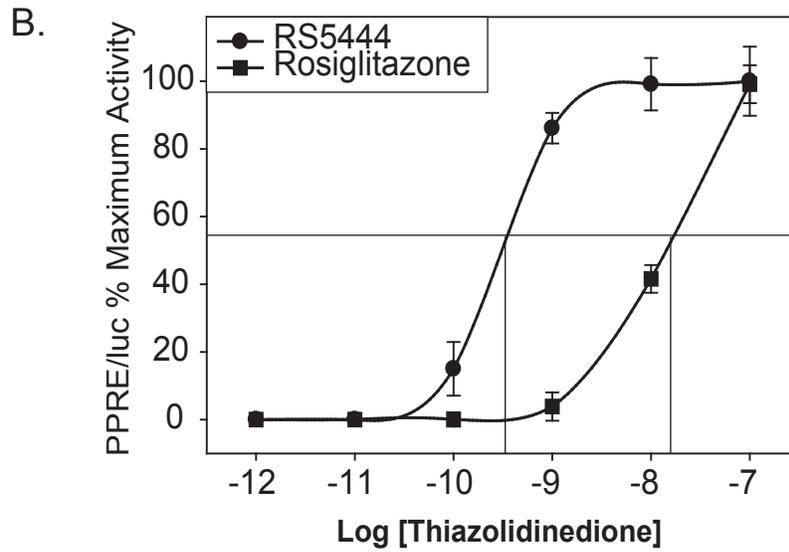
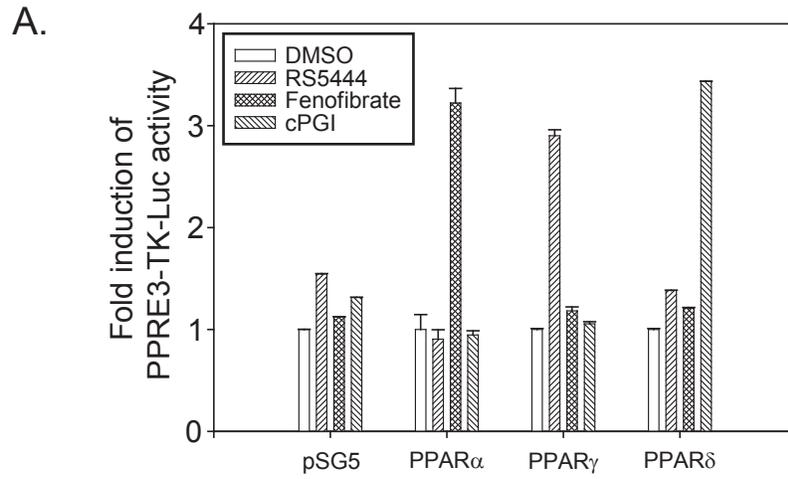


Table 4: Observed fold change of PPAR $\gamma$  marker PDK4 by TZD dose response curve

<b>RS5444 dosage (nM)</b>	<b>MOSER S (PDK4 FC)</b>	<b>RIESP (PDK4 FC)</b>
0	0	0
0.1	5	3.15
1	23.53	22.87
5	33.63	28.58
10	35.57	31.78
50	35.88	32.96
100	35.84	33
<b>Rosiglitazone dosage (nM)</b>	<b>MOSER S (PDK4 FC)</b>	<b>RIESP (PDK4 FC)</b>
0	0	0
1	2.09	2.35
10	13.03	13.41
50	23.22	24.29
100	27.1	27.54
500	30.3	28.24
1000	30.48	28.77

Table 5: Fitting Parameters for Single Binding Affinity Model Assuming Saturation

	MOSER (RS)	RIEP <sup>3</sup> (RS)	MOSER (Rosi)	RIEP <sup>3</sup> (Rosi)
<b>Parameter Values</b>				
Bmax	36.74	33.17	30.9	29.47
Kd	0.56	0.55	14.5	11.26
<b>Std. Errors</b>				
Bmax	0.37	0.73	0.28	0.4
Kd	0.04	0.09	0.74	0.92
<b>95% Confidence Intervals</b>				
Bmax	35.79 to 37.69	31.29 to 35.05	30.19 to 31.61	28.45 to 30.49
Kd	0.46 to 0.66	0.33 to 0.78	12.59 to 16.41	8.90 to 13.62
<b>Goodness of Fit</b>				
Degrees of Freedom	5	5	5	5
R2	0.9996	0.9983	0.9998	0.9994
Residual Sum of Squares	1.9712	7.7715	0.8136	1.8401
Sy.x	0.6279	1.2467	0.4034	0.6067
Fit Status	Converged	Converged	Converged	Converged
<b>Data</b>				
Number of X Values	7	7	7	7
Number of Y Replicates	1	1	1	1
Total Number of Y Values	7	7	7	7
Number of Missing Values	0	0	0	0

## **CHARACTERIZATION OF MOSER S HUMAN COLORECTAL CANCER CELLS**

The MOSER colon carcinoma cell line was established *in vitro* from a spontaneous primary human colon tumor (180). MOSER is an acronym for medium off serum because the initial isolated colonies were identified as incapable of attaching to culture plates in the absence of serum. MOSER S (S for sensitive) cells are sensitive to TGF $\beta$  (181, 182). TGF $\beta$  sensitivity is generally lost in colorectal cancer cells during the transition from adenoma to adenocarcinoma. On this basis, we conclude that MOSER S cell represent a very early adenocarcinoma or late adenoma stage in colon cancer development. MOSER S cells express functional transcripts and proteins of all three PPAR isoforms (182, 183) and are sensitive to TZDs. They display reduced proliferation using saturating amounts of synthetic PPAR $\gamma$  ligand and such effect was reversible by the addition of the PPAR $\gamma$  antagonist GW9662 (183). PPAR $\gamma$  responsiveness is also a marker of differentiated function in intestinal epithelial cells (109), and the observation that MOSER S cells respond to PPAR $\gamma$  agonists is consistent with the conclusion that these cells represent a relatively differentiated, early stage in colon carcinogenesis. Therefore, this cell line offered the best circumstances to study the effects of PPAR $\gamma$  ligands as a potentially exceptional model of mildly transformed colorectal cancer cells.

Although MOSER S cells exhibit many differentiated characteristics of colonic epithelial cells, little was known about the kinds of oncogenic mutations that these cells harbor. To characterize oncogenic mutations within MOSER S cells, genomic DNA was

isolated for sequencing of several common mutations observed in colorectal cancer (184). Specifically, the entire span of exon 15 of the APC gene, exon 3 of human  $\beta$ -catenin, and codons 12, 13, and 61 of human K-Ras were sequenced. An APC mutation at nucleotide position 4069 resulting in a nonsense mutation [**Figure 7A**] was observed. Wild type human APC encodes a 310 kDa protein consisting of 2843 amino acids (185). The MOSER S APC mutation resulted in a truncated protein of 1337 amino acids with a molecular weight of 149.4 kDa [**Figure 7B**]. We detected no mutations to the N-terminus of  $\beta$ -catenin or to codon 12 of K-Ras. On this basis, we conclude that MOSER S cells are mutated for APC function, but do not harbor K-Ras codon 12 mutations. APC mutations are acquired early during transformation of colonic epithelial cells, whereas K-Ras mutations (along with loss of TGF $\beta$  responsiveness) occur during conversion from adenoma to adenocarcinoma (184, 186). These observations are consistent with the conclusion that MOSER S represents a very early stage colon cancer cell.

## **CHARACTERIZATION OF MOSER S CELLULAR PHYSIOLOGY UPON PPAR $\gamma$ ACTIVATION BY RS5444**

### **RS5444 did not alter PPAR $\gamma$ receptor abundance or transcriptional activity**

MOSER S cells were cultured in the presence of 10 nM RS5444 for 24 h. RNA was isolated every 2 h after treatment and protein was isolated every 12 h after treatment. RS5444 rapidly and strongly induced the PPAR $\gamma$  target gene PDK4 [**Figure 8A**]. Error

**Figure 7: Characterization of the APC gene and its expression in MOSER S human colon cancer cells**

**A)** Genomic DNA from MOSER S cells were sequenced for the entire span of exon 15 of the APC gene. A C→T mutation resulted in a non-sense mutation at nucleotide position 4069 (NM\_000038). **B)** 20 µg of total protein lysates from Jurkat, HCT116, and MOSER S cells were resolved using a 6% Tris-Glycine gel. Western blot was used to measure APC and β-actin. Full length and truncated APC protein in these cells were detected by using an antibody that recognizes the N-terminus of wild type APC.

## A.

gi|53759121|ref|NM\_000038.3|

Score = 666 bits (336), Expect = 0.0  
Identities = 341/343 (99%), Gaps = 0/343 (0%)  
Strand=Plus/Plus

Query 19 TCTTTGTCNTCAGCTGAAGATGAAATAGGATGTAATCAGACGACACAGGAAGCAGATTCT 78  
Sbjct 3892 TCTTTGTCATCAGCTGAAGATGAAATAGGATGTAATCAGACGACACAGGAAGCAGATTCT 3951

Query 79 GCTAATACCCTGCAAATAGCAGAAATAAAAGAAAAGATTGGAAGTAGGTCAGCTGAAGAT 138  
Sbjct 3952 GCTAATACCCTGCAAATAGCAGAAATAAAAGAAAAGATTGGAAGTAGGTCAGCTGAAGAT 4011

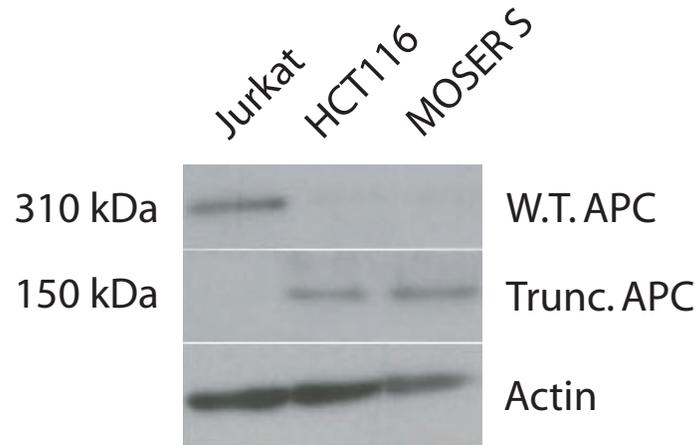
Query 139 CCTGTGAGCGAAGTTCAGCAGTGTACAGCACCCCTAGAACCAAATCCAGCAGACTGTAG 198  
Sbjct 4012 CCTGTGAGCGAAGTTCAGCAGTGTACAGCACCCCTAGAACCAAATCCAGCAGACTGCAG 4071

Query 199 GGTTCTAGTTTATCTTCAGAATCAGCCAGGCACAAAGCTGTTGAATTTTCTTCAGGAGCG 258  
Sbjct 4072 GGTTCTAGTTTATCTTCAGAATCAGCCAGGCACAAAGCTGTTGAATTTTCTTCAGGAGCG 4131

Query 259 AAATCTCCCTCCAAAAGTGGTGCTCAGACACCCAAAAGTCCACCTGAACACTATGTTTCAG 318  
Sbjct 4132 AAATCTCCCTCCAAAAGTGGTGCTCAGACACCCAAAAGTCCACCTGAACACTATGTTTCAG 4191

Query 319 GAGACCCCACTCATGTTTAGCAGATGTACTTCTGTCTCAGTTCAC 361  
Sbjct 4192 GAGACCCCACTCATGTTTAGCAGATGTACTTCTGTCTCAGTTCAC 4234

## B.



bars for hPPAR $\gamma$  are not visible because of the very tight agreement between the three qPCR measurements. RS5444 did not alter the expression of PPAR $\gamma$  mRNA [**Figure 8A**] or protein [**Figure 8B**].

### **Activation of PPAR $\gamma$ caused inhibition of MOSER S anchorage-dependent and – independent growth**

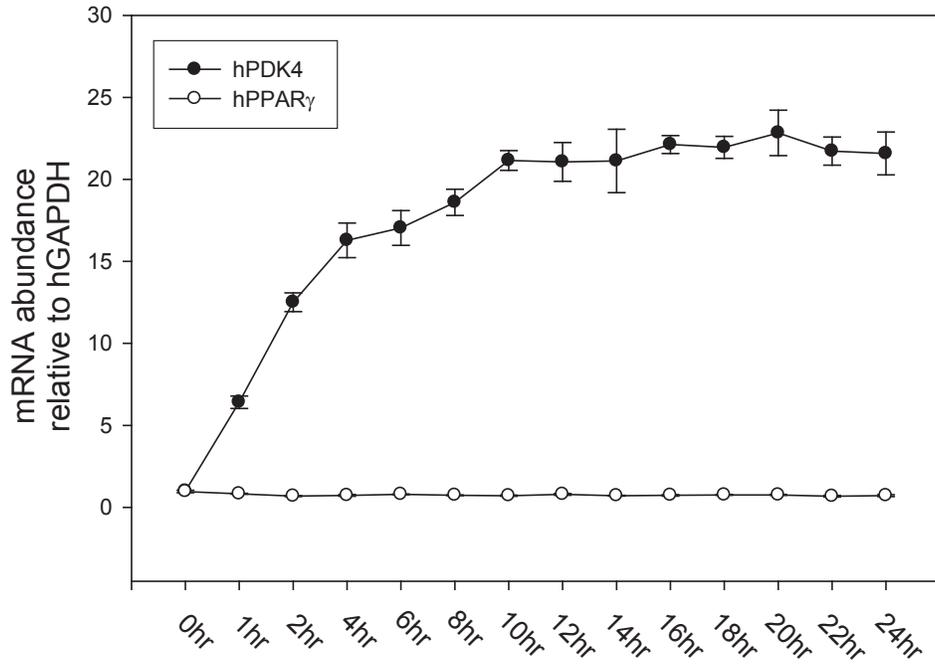
Data from our laboratory indicated that activation of PPAR $\gamma$  inhibited RIE/S3-PPAR $\gamma$  cell proliferation in culture (109). RS5444 and rosiglitazone were also able to inhibit the proliferation of MOSER S cells [**Figure 9**]. Consistent with RS5444 being a much more potent activator of PPAR $\gamma$  transcriptional activity than rosiglitazone [**Figure 6B and 6C**], RS5444 inhibited MOSER S cell proliferation at a much lower concentration (10 nM) than rosiglitazone (1  $\mu$ M).

MOSER S cells are transformed colonic epithelial cells and can grow in soft agar in anchorage-independent manner [**Figure 10**]. RS5444 significantly suppressed the ability of MOSER S cell to grow on soft agar. Although the method employed to analyze the suppression of colony growth did not take into account the amount of colony size reduction, RS5444 appeared to reduce the size of colonies and not the number of colonies (data not shown). The number of colonies at the time of re-plating, after treatment with RS5444, was the same. These data indicate that PPAR $\gamma$  inhibits growth of MOSER S cells both on plastic (anchorage-dependent) and in soft agar (anchorage-independent).

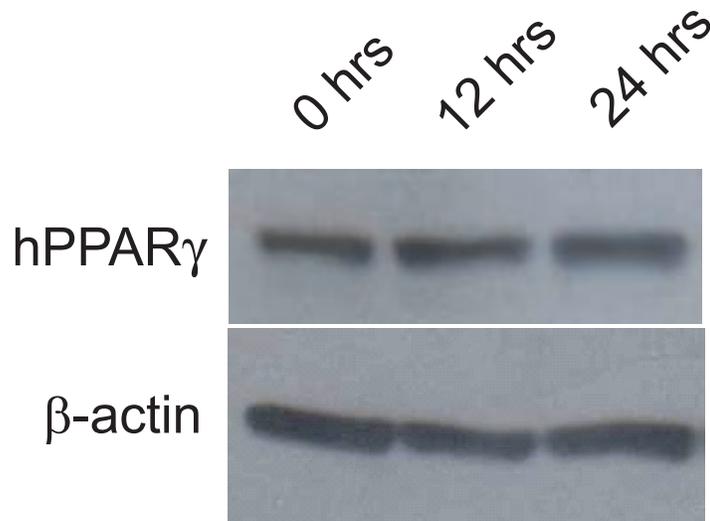
**Figure 8: *RS5444 did not alter PPAR $\gamma$  expression level or transcriptional activity in MOSER S cells***

MOSER S cells were treated with 10 nM RS5444 for 24 h. **A)** Total RNA was isolated every 2 h following RS5444 treatment. The abundance of PDK4 and PPAR $\gamma$  transcript were measured by qPCR and normalized to that of GAPDH. Data represent the mean  $\pm$  SD, n = 3. **B)** 10  $\mu$ g of total protein lysates isolated 0, 12, and 24 h after RS5444 treatment were resolved using a 10% Tris-Glycine gel. Western blot was used to measure PPAR $\gamma$  and  $\beta$ -actin

A.

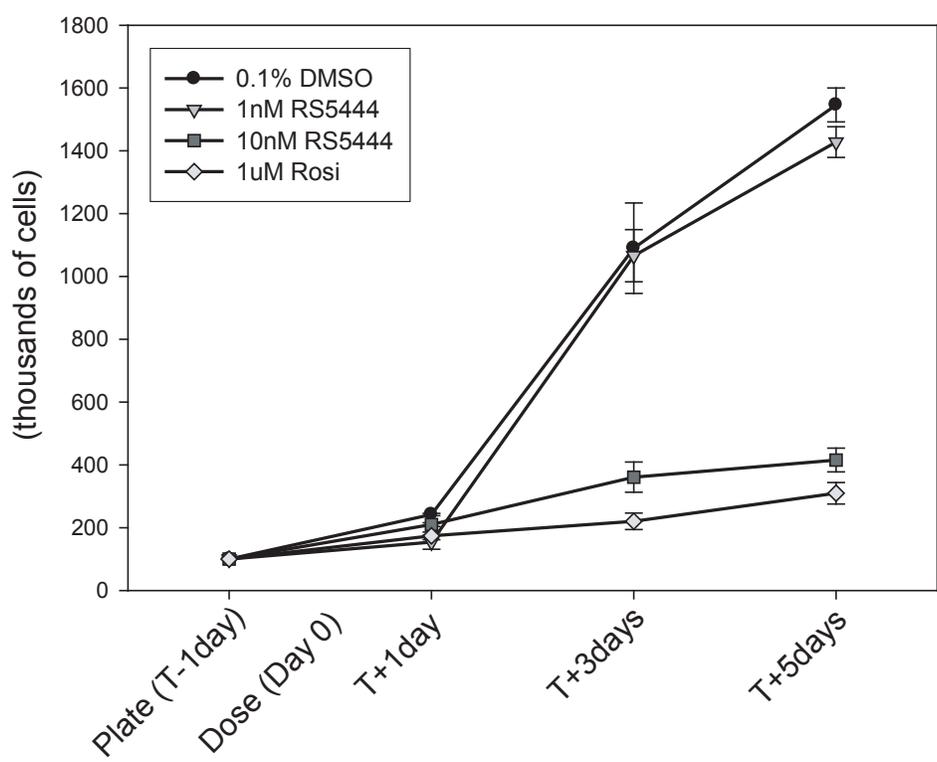


B.



**Figure 9: *RS5444 inhibited MOSER S cell proliferation under anchorage-dependent conditions***

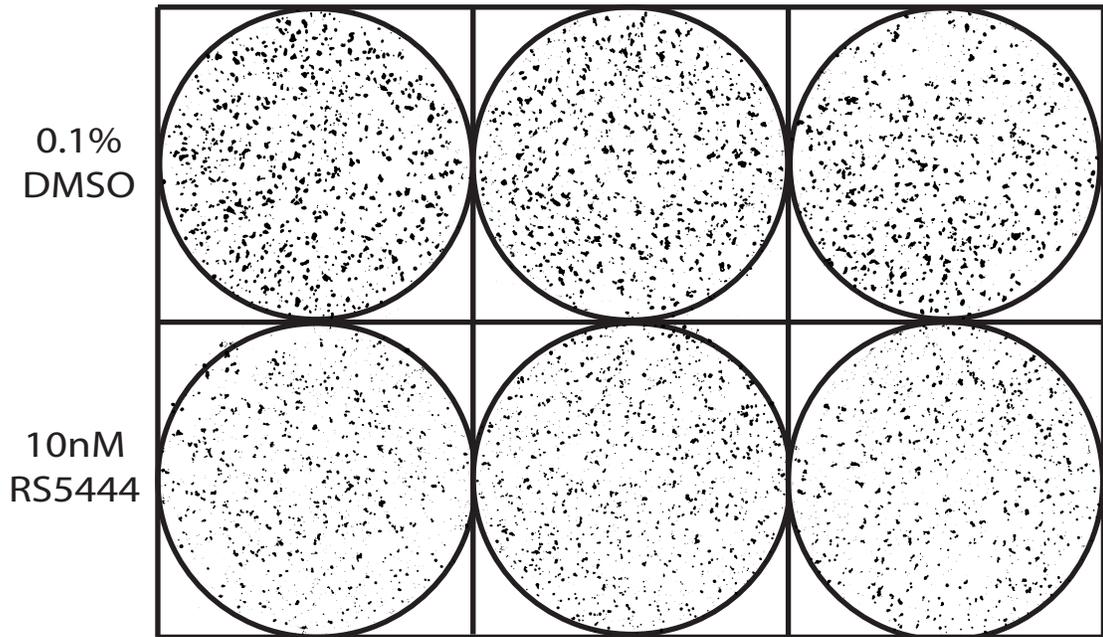
MOSER S cells were plated in 60 mm dishes at  $1 \times 10^5$  cells per dish. Cells were treated with 0.1% DMSO, 1 or 10 nM RS5444, or 1  $\mu$ M rosiglitazone. Cell proliferation was measured by Beckman Coulter counter on 0, 1, 3 and 5 days after treatment. Data represent the mean  $\pm$  SD, n = 3



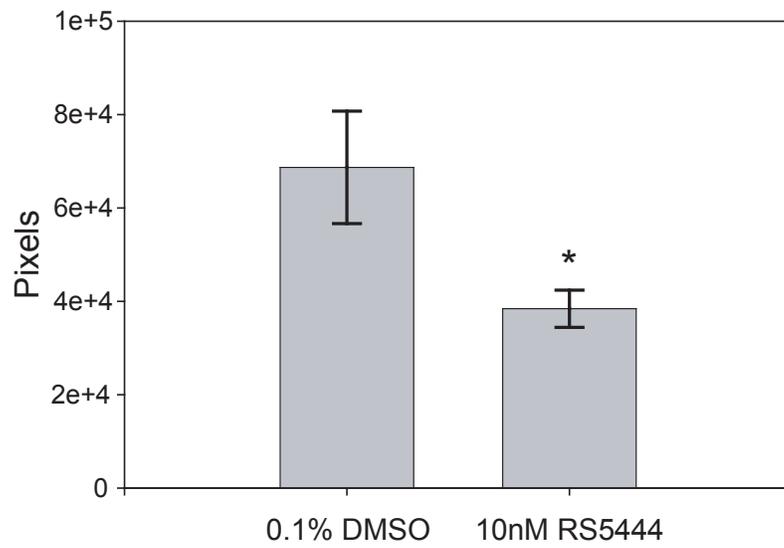
**Figure 10: *RS5444 inhibited MOSER S cell proliferation in soft agar***

**A)**  $1 \times 10^5$  MOSER S cells suspended in 0.75% soft agar containing 0.1% DMSO or 10 nM RS5444 were allowed to grow for 2 weeks in 60mm culture plates. The cultures were fixed and stained with Giemsa staining solution. The colonies were photographed with a Kodak 4800 photographic system using an EL Logic 100 camera. Representative colonies are shown. **B)** The number of colonies was counted as described in Material and Methods. Data represent the mean  $\pm$  SD, n = 6. \*, p < 0.002

A.



B.



### ***The anti-proliferative effect of RS5444 was irreversible***

MOSER S cells were treated with 0.1% DMSO or 10 nM RS5444 for 24 hours and then re-plated in media that did not contain RS5444. Cells that can still proliferate will form colonies. If the anti-proliferative effect of RS4444 were irreversible, one would expect that cells pre-treated with RS5444 would not proliferate and would not form colonies. As shown in **Figure 11**, the plating efficiency of RS5444 pre-treated cells was less than 60% of that of control cells. These data suggest that activation of PPAR $\gamma$  in MOSER S cells by RS5444 caused a subset of cells to permanently stop proliferation. Irreversible withdrawal from the cell cycle is one of the hallmarks of differentiation, and our data suggest that PPAR $\gamma$  may cause differentiation of some MOSER S cells. We did not investigate this issue further and it is clear that additional work will be required to explain the precise mechanisms of action responsible for these observations.

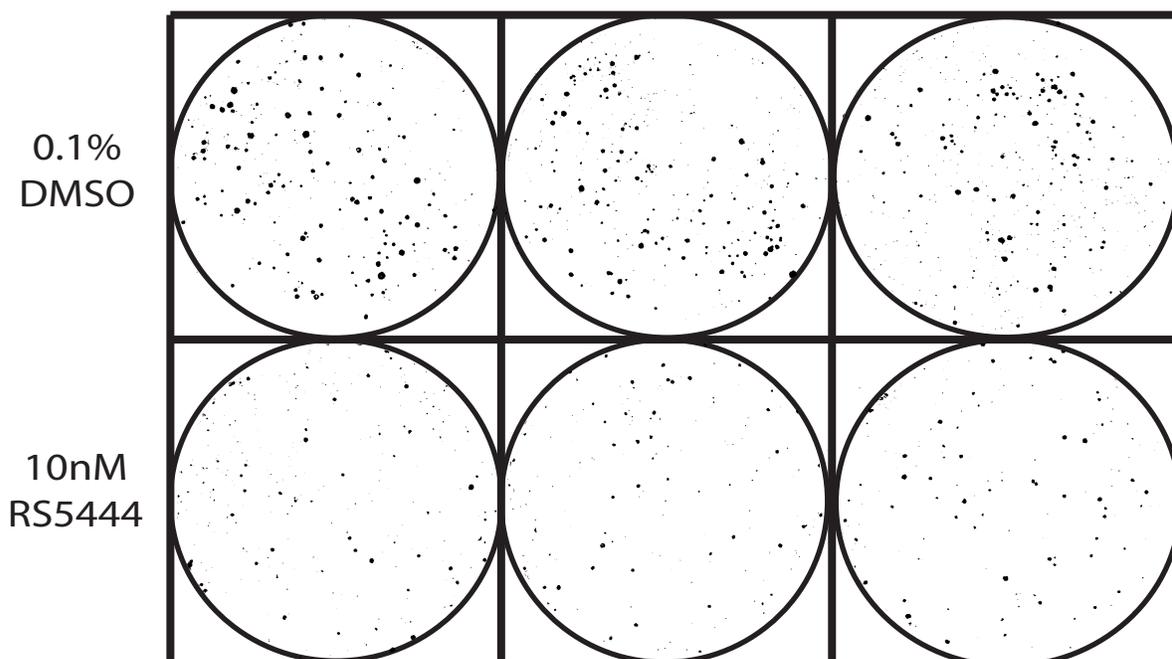
### **Activation of PPAR $\gamma$ by RS5444 inhibited invasiveness, but not motility, of MOSER S cells**

To test the effect of RS5444 on MOSER S cell invasion and migration, cells pre-treated with 0.1% DMSO or 10 nM RS5444 for 24 h were plated into matrigel-coated transwells to measure invasion. Collagen-coated transwells were used to measure migration. Invasive cells would digest the extracellular matrix, migrate through the matrigel layer and the chamber membrane, and attach to the outside surface of the membrane. On the other hand, migration (or motility) only measures cell's ability to migrate through the collagen layer. As shown in **Figure 12**, MOSER S cells treated with

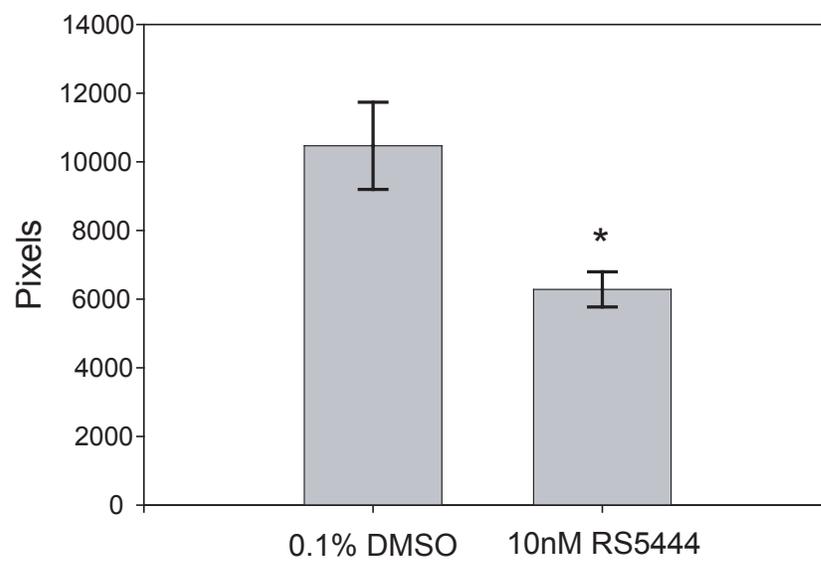
**Figure 11: *RS5444's anti-proliferative effect was irreversible***

**A)** MOSER S cells were treated with 0.1% DMSO or 10 nM RS5444 for 4 days. Cultures were then replated at  $1 \times 10^4$  cells per 60 mm dish and maintained in media that did not contain RS5444 for 10 days. Colonies were stained with Giemsa staining solution. The colonies were photographed with a Kodak 4800 photographic system using an EL Logic 100 camera. Representative colonies are shown. **B)** The number of colonies was counted using the method described in Material and Methods. Data represent the mean  $\pm$  SD, n = 6. \*, p < 0.0005.

A.

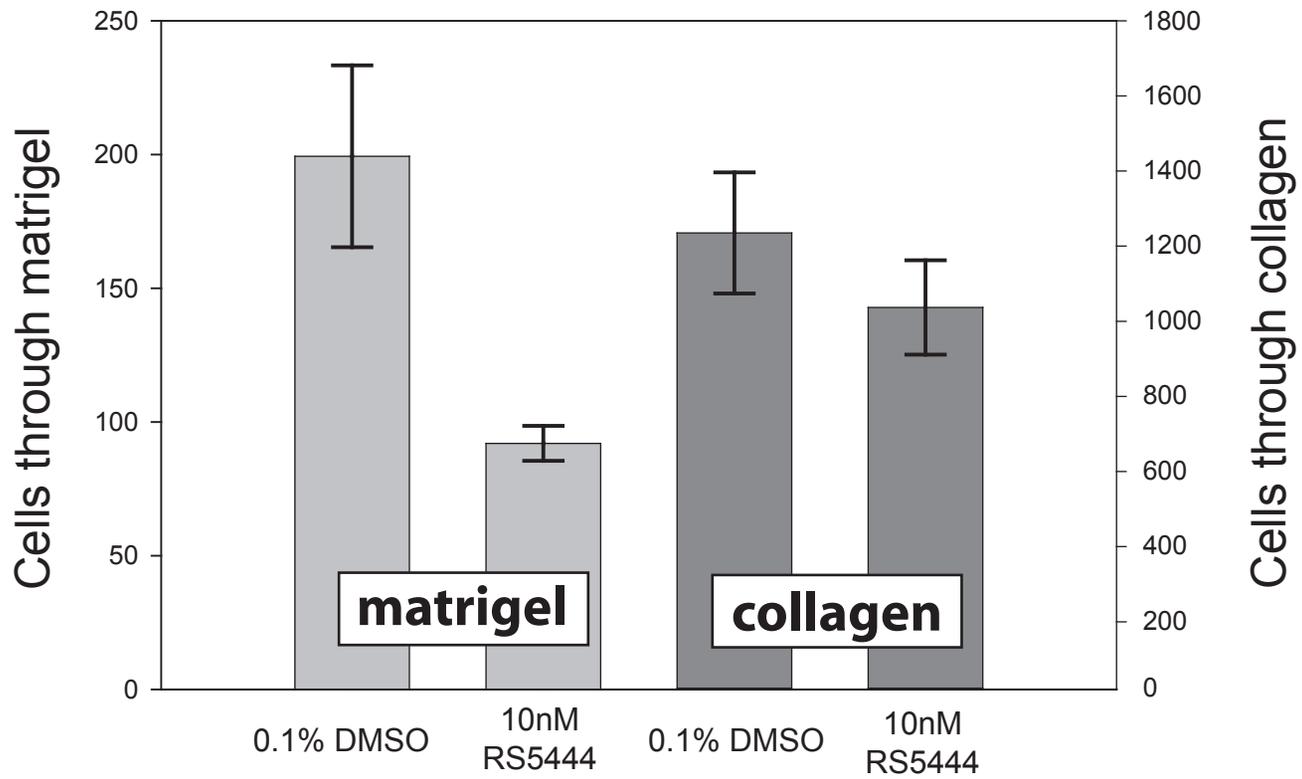


B.



**Figure 12: *Activation of PPAR $\gamma$  by RS5444 inhibited invasion, but not migration, of MOSER S cells***

MOSER S cells were treated with 0.1% DMSO or 10 nM RS5444 for 24 h.  $5 \times 10^4$  pre-treated cells were then plated in matrigel coated transwells for invasion assay or collagen coated transwells for migration assay. Cells attaching to the outside surface of the transwells were stained with crystal violet and counted under reverse microscope. Data represent the mean  $\pm$  SD, n = 3. \*, p < 0.01.



RS5444 were significant less invasive than control cells. In contrast, RS5444 did not alter MOSER S cell motility in the migration assay. These data suggest that activation of PPAR $\gamma$  suppresses MOSER S cell invasiveness by inhibiting the ability of MOSER S cells to digest the extracellular matrix, rather than by inhibiting the motility of these cells.

## **FUNCTIONAL GENOMICS ANALYSIS OF RS5444 EFFECT ON MOSER S COLORECTAL CANCER CELLS**

The characterization experiments gave a clear, albeit high level, picture of the kinds of processes that are affected by activated PPAR $\gamma$ . Specifically, these data indicate that PPAR $\gamma$  regulates genomic functions that are associated with proliferation and transformed properties, including anchorage-independent growth and invasion. To gain mechanistic insight into these responses, it was important to obtain the best possible functional genomic dataset covering the period of time when MOSER S cells display maximal influence of PPAR $\gamma$ -mediated pathway regulation: during the first twenty-four hours after addition of the PPAR $\gamma$  agonist RS5444. We chose a time course experimental design to capture the transcript kinetics within this 24hr time window.

## EXPERIMENTAL DESIGN OF MICROARRAY TIME COURSE

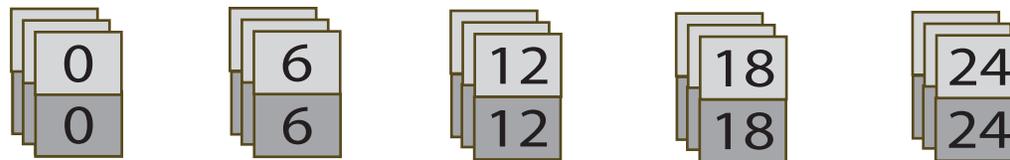
### EXPERIMENT

The Affymetrix HGU-133plus2 microarray was used for all microarray experiments. The HGU-133plus2 form factor was chosen because the entire human genome is represented on a single chip. This chip contains 54,675 probe sets measuring the expression level of more than 47,000 transcripts and their variants, including approximately 38,500 well-characterized genes.

The experimental design of the two-class time course experiment is illustrated in **Figure 13**. The set of light grey blocks represents three replicate 60 mm culture dishes containing MOSER S cells cultured in the presence of 10 nM RS5444. The set of dark grey blocks represents three replicate 60 mm culture dishes containing MOSER S cells cultured in the absence of RS5444 (presence of 0.1% DMSO solvent). Each plate was initially seeded with 300,000 cells per 60 mm plate and prepared for experimentation as discussed in Materials and Methods. The numbers denote the time in hours at which total RNA would be isolated. In practice, four technical replicates were performed for each class at each time point. This was done as a precaution in case one of the replicates failed to pass our quality control regime (see below).

**Figure 13: *Experimental design of microarray time course***

Triplicate mRNA samples were isolated from culture and prepared for Affymetrix HGU-133plus2 microarrays at 0 h, 6 h, 12 h, 18h, and 24 h. mRNA from samples cultured in the presence of 10 nM RS4444 (top row, light gray) were isolated at the times indicated. mRNA from samples cultured in the presence of 0.1% DMSO (bottom row, dark gray ) were isolated every six hours at the times indicated



 DMSO

 RS5444

time indicated is in hours

## QUALITY CONTROL AND PREPROCESSING OF AFFYMETRIX

### MICROARRAY DATA

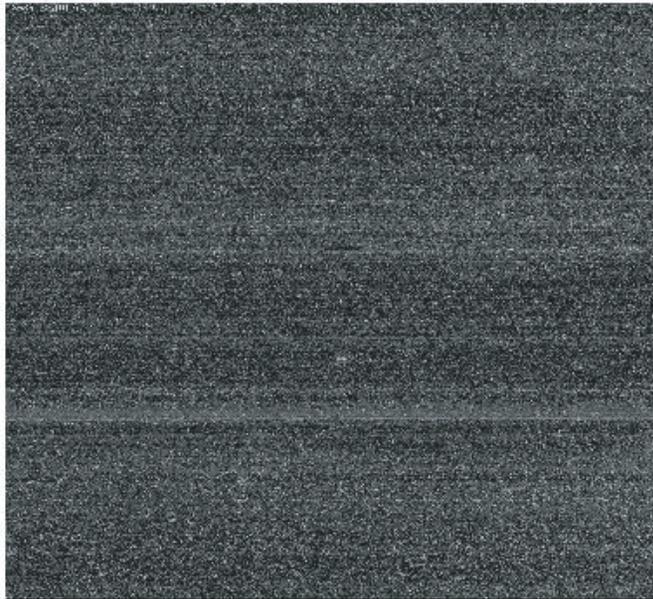
#### Quality Control Step: Visual Inspection of CEL images

An example of both Affymetrix .CEL image files that failed and passed visual inspection are illustrated in **Figure 14**. The CEL file image 20-2638.cel depicts a microarray data set that did not pass inspection. The ghost-like letter “D” is most likely a manufacturing error, however it is impossible to know the exact cause. The highlighted regions forming this artifact would almost certainly bias probe-set expression values, and therefore this replicate was rejected. Of all 54 microarrays processed by the UTMB genomics core, only 7 displayed ghosting, scratches or other artifacts that required repeating. The genomics core was given enough RNA to repeat any replicate of any time point at least once for such contingencies. The CEL file image 1-4G638.cel represents the majority of CEL images passing visual inspection. The characteristic snow field

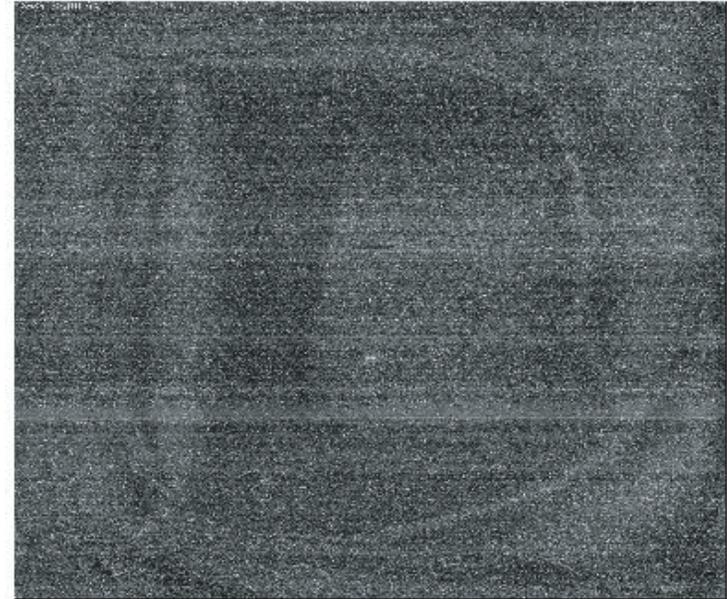
**Figure 14: *Inspection of microarray images for artifacts***

Examples of Affymetrix CEL images. CEL file 1-4G638 represents a replicate that passed visual inspection, where no clear artifact that may bias the analysis can be found. CEL file 20-2G638 represents a replicate that did not pass visual inspection.

1-4G638.CEL



20-2G638.CEL



represents a properly manufactured microarray chip and proper liquid handling of cRNA solution during the hybridization phase of preparation.

A note on cell file nomenclature: The nomenclature X-Y[R]-GZZZ.cel denotes information on the time point, the technical replicate, whether it was repeated at the UTMB molecular core, and the molecular core's run for each class (plus and minus RS5444). X denotes the time point. Y denotes the technical replicate. R denotes a replicate had to be repeated by the molecular core. GZZZ denotes the batch run for the particular class. G638 was the batch process for all replicates in the presence of RS5444. G659 was the batch process for all replicates in the absence of RS5444.

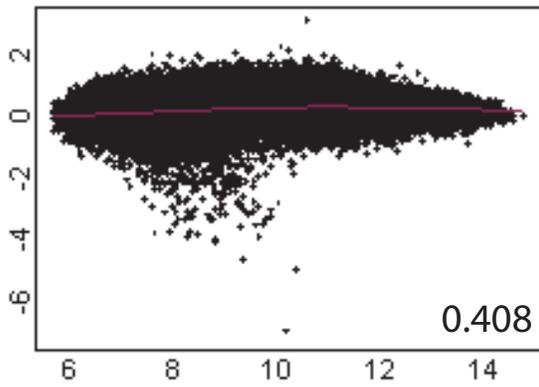
### **Determination of Preprocessing Algorithm**

Once the CEL files that passed visual inspection were identified, the next step was to determine the proper preprocessing algorithm to expression values. Three different algorithms were considered and are discussed in more detail in Materials and Methods: Affymetrix's MAS5, Bioconductor's RMA, and Bioconductor's GCRMA. Bland-Altman plots (a.k.a. MvA plots) generated in Splus' Array Analyzer were used to compare the performance of each test versus raw, unprocessed data [**Figure 15**]. For space considerations, a comparison of the second and third 24hr time points in the absence of RS5444 is shown.

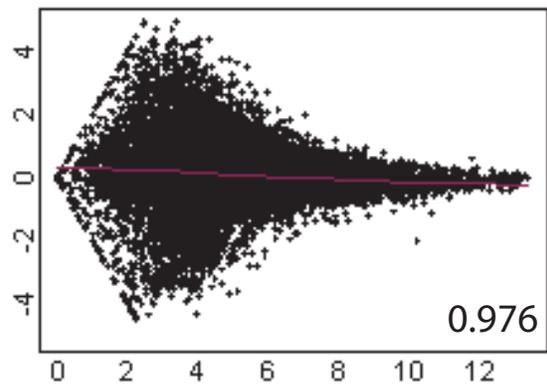
It can be seen from this small sampling that GCRMA provides a tighter, more uniform processing of expression values, with the lowest overall lowess regression score

**Figure 15: *Bland-Altman plot of preprocessing methods***

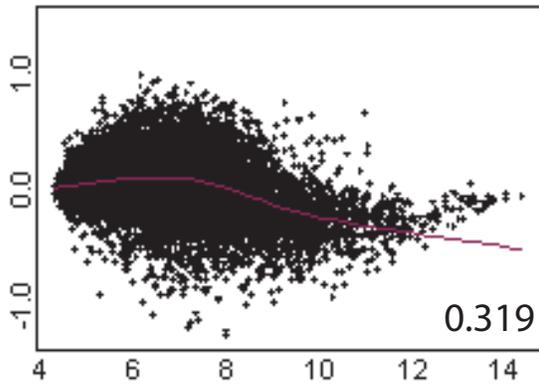
Bland-Altman plot of the second and third control replicate at the 24 h time point. The GCRMA preprocessing method outperforms both MAS5 and RMA (lower lowess scores are better).



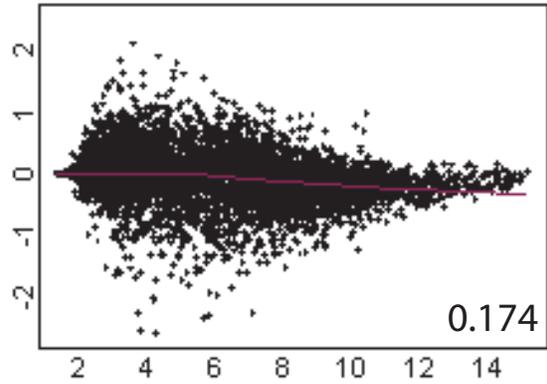
Unprocessed



MAS5



RMA



GCRMA

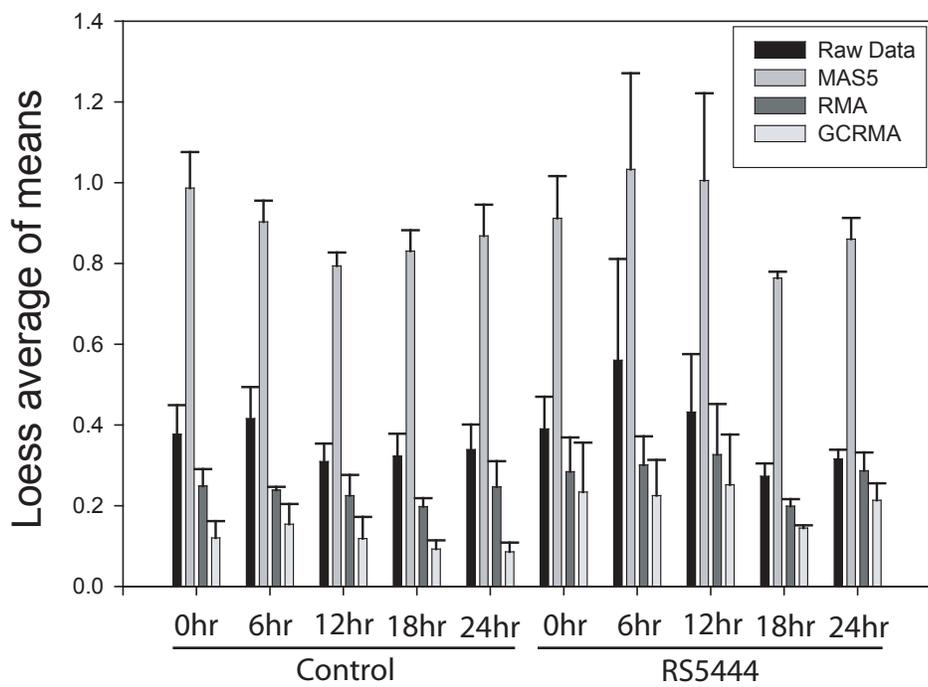
of 0.174. Lower scores are better. Perfect agreement between two samples would result in a lowess score of zero. RMA preprocessing resulted in a lowess score of 0.319. Note the odd skewing of values at higher means that was observed when summarization was carried out with RMA. This was characteristic of many pair-wise RMA plots. MAS5 (lowess score of 0.976) fared worse than unprocessed data (lowess score of 0.408). A full list of lowess scores from the MvA plots is provided for all chips in **Table 6**. Averaging across all lowess scores gives an average score of 0.374 for raw unprocessed data, 0.896 for MAS5 preprocessed data, 0.255 for RMA preprocessed data, and 0.164 for GCRMA preprocessed data. This table is summarized graphically in **Figure 16**.

Table 6: Lowess scores of preprocessing methods

Pairwise set	Raw Data	MAS5	RMA	GCRMA
<b>0hr Control</b>				
rep 1 vs rep 2	0.363	0.978	0.241	0.0869
rep 2 vs rep 3	0.313	0.902	0.211	0.107
rep 1 vs rep 3	0.455	1.08	0.294	0.167
<b>6hr Control</b>				
rep 1 vs rep 2	0.499	0.959	0.24	0.103
rep 2 vs rep 3	0.345	0.855	0.23	0.203
rep 1 vs rep 3	0.404	0.895	0.246	0.157
<b>12hr Control</b>				
rep 1 vs rep 2	0.333	0.827	0.254	0.152
rep 2 vs rep 3	0.258	0.76	0.166	0.0569
rep 1 vs rep 3	0.337	0.794	0.255	0.147
<b>18hr Control</b>				
rep 1 vs rep 2	0.366	0.89	0.205	0.108
rep 2 vs rep 3	0.26	0.803	0.175	0.102
rep 1 vs rep 3	0.342	0.799	0.214	0.0683
<b>24hr Control</b>				
rep 1 vs rep 2	0.317	0.871	0.243	0.0792
rep 2 vs rep 3	0.409	0.944	0.312	0.111
rep 1 vs rep 3	0.291	0.79	0.185	0.0668
<b>0hr Treatment</b>				
rep 1 vs rep 2	0.417	0.904	0.278	0.214
rep 2 vs rep 3	0.301	0.811	0.203	0.122
rep 1 vs rep 3	0.453	1.02	0.372	0.365
<b>6hr Treatment</b>				
rep 1 vs rep 2	0.662	1.15	0.31	0.215
rep 2 vs rep 3	0.275	0.759	0.226	0.142
rep 1 vs rep 3	0.744	1.19	0.367	0.318
<b>12hr Treatment</b>				
rep 1 vs rep 2	0.526	1.13	0.387	0.308
rep 2 vs rep 3	0.503	1.13	0.41	0.338
rep 1 vs rep 3	0.267	0.756	0.183	0.109
<b>18hr Treatment</b>				
rep 1 vs rep 2	0.296	0.755	0.215	0.139
rep 2 vs rep 3	0.286	0.782	0.201	0.152
rep 1 vs rep 3	0.237	0.755	0.181	0.143
<b>24hr Treatment</b>				
rep 1 vs rep 2	0.338	0.916	0.338	0.251
rep 2 vs rep 3	0.317	0.812	0.254	0.168
rep 1 vs rep 3	0.293	0.853	0.268	0.221
<b>AVERAGE</b>	<b>0.374</b>	<b>0.896</b>	<b>0.255</b>	<b>0.164</b>

**Figure 16: *Averages of lowess regression scores at each time point***

GCRMA preprocessing outperforms both MAS5 and RMA. Data denotes averages of lowess scores by comparing the 1<sup>st</sup> replicate versus 2<sup>nd</sup> replicate, the 2<sup>nd</sup> replicate versus the 3<sup>rd</sup> replicate, and the 1<sup>st</sup> replicate versus the 3<sup>rd</sup> replicate. (n = 3, mean  $\pm$  SD).



### **Quality Control Step: Unsupervised Clustering**

Microarray two-class time series data were preprocessed by GCRMA summarization and analyzed by principal components analysis using Splus' Array Analyzer [Figure 17]. Plotting each microarray chip along its first and second principal components shows good separation between classes (control versus treatment). Likewise, while the control class clusters irrespective of time, the treatment class clusters with early time points near the controls and later time points increasing in distance as a function of time. This clustering pattern suggests tight regulation of the MOSER S genome in the presence of RS5444, while genome activity in the absence of drug follows some latent, endogenous program. These results are exactly what one would expect to see following a proper longitudinal two-class microarray experiment where control data was clustered with data depicting a biological perturbation over some finite time span.

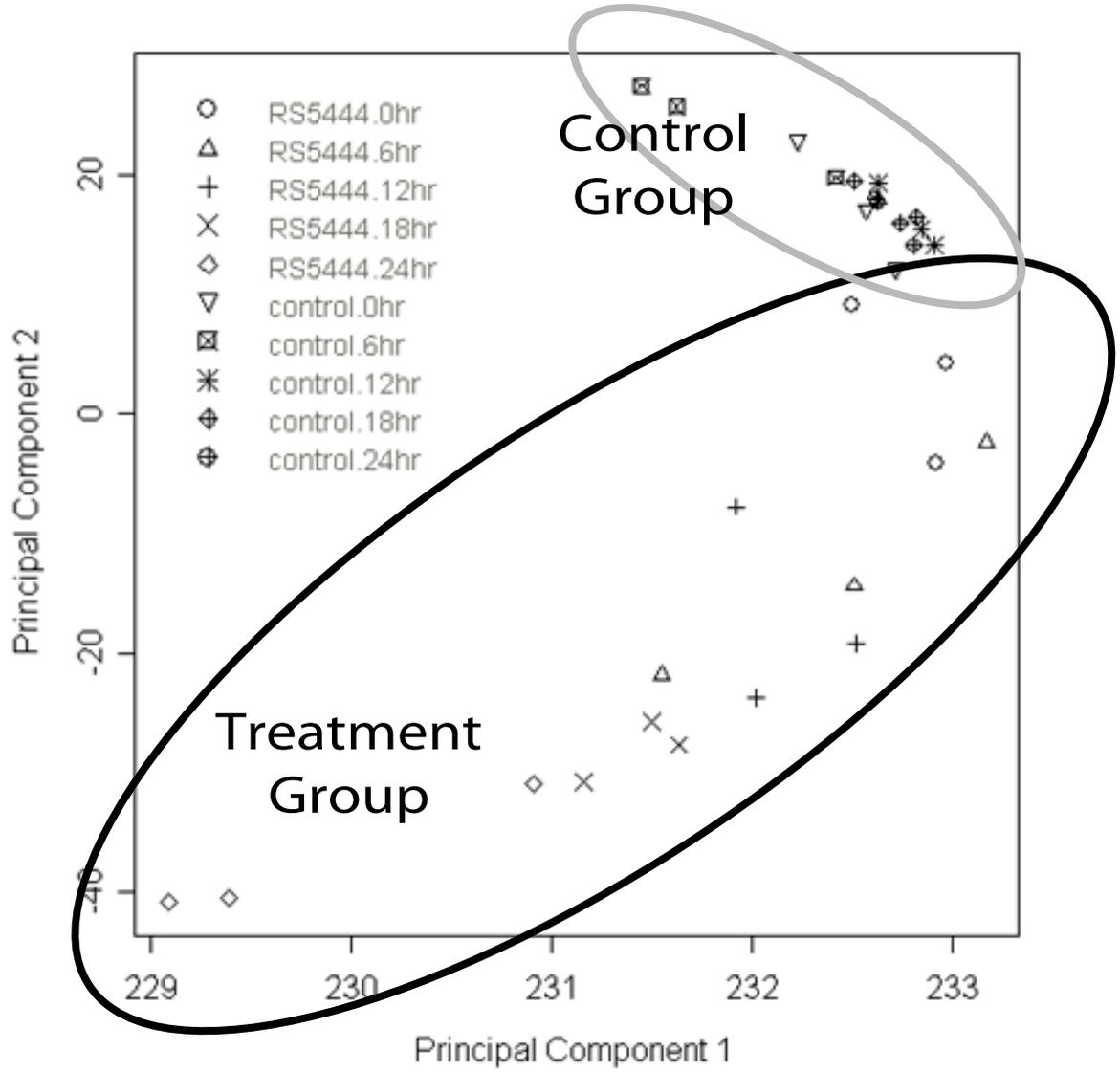
### **Probe Set Filters**

Affymetrix MAS5 change calls were computed as described in Materials and Methods. Of the 54,675 probe sets on the HGU-133plus2 Affymetrix human array, excluding housekeeping probe sets, 10,227 probe sets had a change call of present "P" or marginally present "M" across all time points, over all replicates, within both conditions. Only this reduced set of probe sets was used for further analysis. Implementing the change call filter guarantees analysis results of only those genes considered "on" over the time span of the experiment in the biological context, whether differentially expressed or not.

**Figure 17: *Principal Components Analysis (PCA) of microarray time course***

GCRMA summarization data (log base 2) was used as input for PCA analysis. The first two principle components are plotted. Control replicates are grouped together (gray oval). Treated replicates are grouped together (black oval).

# Principal Component Plot



## **Analysis of Performance of Differential Expression Models**

Three computational models were examined to determine their statistical power based on a predefined group of both positive and negative controls. These models employed, in order of decreasing complexity, an empirical Bayes machine learning algorithm, a polynomial regression algorithm, and a non-parametric algorithm. For these experiments the two-class longitudinal dataset format was used. The performance scoring algorithm boot-straps a candidate subset of data from the original two-class longitudinal dataset and, along with the positive and negative controls, and scores each probe set by the metric defined by each of the three differential expression models. In this context, boot-strapping refers to the measurement of a test using data in which the best test will be used to analyze the data. Boot-strap statistics were chosen over an independent verification experiment for cost considerations. Indeed, under ideal circumstances, separate time course data would be used to verify the results shown. However, obtaining these data is cost-prohibitive.

Two different figures of merit were used for the boot-strap test. For the non-parametric and polynomial regression model, this metric is an adjusted p-value. For the empirical Bayes model, this metric is the MB-statistic (see below). The boot-strap algorithm runs these tests five thousand times, each time picking a new random subset of probe sets from the two-class longitudinal dataset, seeding in the controls, and scoring with the three different models. The scores are tabulated with each run. The final result lists each of the controls and a percentage metric for each of the three models. The percentage represents the number of times a particular model determined a control as

significant (i.e. # of times significant / # of total tests). This percentage can be defined as the model's statistical power for a particular control. Perfect statistical power for a particular positive control and a particular model would be defined as 100% power, where the model determined the positive control significant over all iterations of the boot-strapping performance test.

### ***Structure of input data***

The input data's structure mimics the two-class longitudinal dataset. That is, for every time point (0 h, 6 h 12 h, 18 h, and 24 h), there are triplicate replicates from the control set (in the absence of RS5444) and triplicate replicates from the treated set (in the presence of RS5444). Sixteen positive controls and sixteen negative controls were chosen for this experiment. Positive controls and negative controls were selected based on a subjective assessment of their regulation. Probe sets representing varying degrees of regulation strength were chosen as positive controls. Likewise, probe sets displaying no regulation whatsoever were chosen as negative controls.

### ***Pseudo-code of algorithm***

1. For 1 through 5000 iterations
  - a. Pick 1468 probe sets randomly (+/- RS5444 @ 0, 6, 12, 18, and 24 h)
  - b. Seed in 32 positive and negative controls (total 1500)
  - c. Calculate differential expression metric for all probe sets plus controls

- i. Invoke non-parametric algorithm with dataset
  - ii. Invoke polynomial regression algorithm with dataset
  - iii. Invoke empirical Bayes algorithm with dataset
  - iv. Tabulate running percentage of true positives for each control
2. End For loop

**Table 7** lists the statistical power of the models defined by the boot-strapping test for both the positive and negative controls. For the positive controls, higher power is better. Power for this test is defined as the average percentage where a model chose a true positive. The empirical Bayes HotellingT2 algorithm had an average statistical power of ~80%. The polynomial regression algorithm had an average statistical power of ~69%. The non-parametric algorithm had an average statistical power of ~54%. Under these test conditions, the empirical Bayes model clearly outperforms the other two models.

For the negative controls, lower power is better. Power for this test is defined as the average percentage where a model did not choose a false positive. The empirical Bayes HotellingT2 algorithm had an average statistical power of 0.0%. The polynomial regression algorithm had an average statistical power of 0.009%. The non-parametric algorithm had an average statistical power of 0.02%. While each of the three models was good at picking true negatives (i.e. not picking false positives), the empirical Bayes model again scored better than the other two.

Table 7: Statistical Power of Differential Expression Models by Boot-Strapping with Replacement Test

Statistical power by positive controls (n = 1500 probe sets)

Probe Set	Gene Symbol	Empirical Bayes	Polynomial Regression	Non Parametric
222912_at	ARRB1	5000/5000 = 100.00%	5000/5000 = 100.00%	5000/5000 = 100.00%
228250_at	KIAA1961	5000/5000 = 100.00%	5000/5000 = 100.00%	5000/5000 = 100.00%
209498_at	CEACAM1	5000/5000 = 100.00%	5000/5000 = 100.00%	5000/5000 = 100.00%
221276_s_at	SYNC1	5000/5000 = 100.00%	5000/5000 = 100.00%	5000/5000 = 100.00%
205471_s_at	DACH1	5000/5000 = 100.00%	5000/5000 = 100.00%	5000/5000 = 100.00%
201470_at	GSTO1	5000/5000 = 100.00%	5000/5000 = 100.00%	5000/5000 = 100.00%
222787_s_at	FLJ11273	5000/5000 = 100.00%	698/5000 = 13.92%	5000/5000 = 100.00%
223059_s_at	C10orf45	5000/5000 = 100.00%	5000/5000 = 100.00%	4864/5000 = 97.28%
213357_at	GTF2H5	5000/5000 = 100.00%	5000/5000 = 100.00%	9/5000 = 0.18%
225239_at	---	5000/5000 = 100.00%	4962/5000 = 99.2%	1925/5000 = 38.50%
218117_at	RBX1	4971/5000 = 99.42%	3593/5000 = 71.86%	832/5000 = 16.64%
202299_s_at	HBXIP	4749/5000 = 94.98%	2836/5000 = 56.72%	323/5000 = 6.4%
231174_s_at	EPB41L2	3235/5000 = 64.70%	1769/5000 = 35.38%	187/5000 = 3.74%
209205_s_at	LMO4	1022/5000 = 20.44%	599/5000 = 11.98%	0/5000 = 0.00%
202760_s_at	AKAP2	12/5000 = 0.24%	333/5000 = 0.07%	0/5000 = 0.00%
211318_s_at	RAE1	0/5000 = 0.00%	164/5000 = 0.03%	0/5000 = 0.00%
<b>AVERAGES</b>		<b>79.99%</b>	<b>68.69%</b>	<b>53.93%</b>

Statistical power by negative controls (n = 1500 probe sets)

Probe Set	Gene Symbol	Empirical Bayes	Polynomial Regression	Non Parametric
1562861_at	---	0/5000 = 0.00%	0/5000 = 0.00%	0/5000 = 0.00%
1555820_a_at	FLJ20345	0/5000 = 0.00%	0/5000 = 0.00%	0/5000 = 0.00%
48612_at	N4BP1	0/5000 = 0.00%	584/5000 = 11.68%	5/5000 = 0.001%
239718_at	---	0/5000 = 0.00%	0/5000 = 0.00%	0/5000 = 0.00%
1555843_at	HNRPM	0/5000 = 0.00%	0/5000 = 0.00%	894/5000 = 17.88%
228528_at	---	0/5000 = 0.00%	0/5000 = 0.00%	0/5000 = 0.00%
1554344_s_at	AQP12B	0/5000 = 0.00%	0/5000 = 0.00%	0/5000 = 0.00%
1562256_at	NALP1	0/5000 = 0.00%	0/5000 = 0.00%	0/5000 = 0.00%
207105_s_at	PIK3R2	0/5000 = 0.00%	0/5000 = 0.00%	0/5000 = 0.00%
1562257_x_at	NALP1	0/5000 = 0.00%	0/5000 = 0.00%	0/5000 = 0.00%
37793_r_at	RAD51L3	0/5000 = 0.00%	0/5000 = 0.00%	532/5000 = 10.64%
1557818_x_at	---	0/5000 = 0.00%	111/5000 = 2.22%	0/5000 = 0.00%
1559048_at	KIAA1447	0/5000 = 0.00%	0/5000 = 0.00%	0/5000 = 0.00%
49878_at	PEX16	0/5000 = 0.00%	0/5000 = 0.00%	162/5000 = 3.24%
60528_at	PLA2G4B	0/5000 = 0.00%	0/5000 = 0.00%	0/5000 = 0.00%
31861_at	IGHMBP2	0/5000 = 0.00%	0/5000 = 0.00%	0/5000 = 0.00%
<b>AVERAGES</b>		<b>0.00%</b>	<b>0.87%</b>	<b>1.99%</b>

## **Differential Expression Analysis by empirical Bayes HotellingT2 score**

Given the results of the boot-strapping statistical power test, the statistical machine learning algorithm employing an empirical Bayesian implementation of the multivariate HotellingT2 scores was chosen to define differentially expressed genes over time. As input, the model was fed triplicate GCRMA expression values for probe sets at time points 0 h, 6 h, 12 h, 18 h, and 24 h in both the presence and absence of RS5444. Only the 10,277 probe sets that had MAS5 change call of “P” or “M” across all conditions were included. The control and treatment replicates were normalized such that a control’s median expression value at time point 0 h was equal to the treatment’s median expression value at time point 0 h. Subsequent control time points, for each replicate, were adjusted by this distance. For example, if the distance between the median of treatment and control groups at time zero, for a particular probe set, was 1.0, then each subsequent treatment measurements for that probe set were adjusted by this value. This adjustment guarantees that both control and treatment classes’ kinetics begin at the same expression value.

The empirical Bayes HotellingT2 results can be treated as a monotonically decreasing list. That is, each HotellingT2 score is unique for a particular probe set. The list of scores can then be force ranked so that the probe set with the highest score is first and the probe set with the lowest score is last. Likewise, because the model uses Bayesian statistics, each score has an associated posterior odds score. This score is defined by Tai and Speed as:

$$O = \frac{p}{1-p} \frac{P(\tilde{t}|I=1)}{P(\tilde{t}|I=0)}$$

where the posterior odds  $O$ , is equal to the probability that the expected timecourse  $\mu$  is nonconstant (i.e.  $I=1$ ) divided by the probability that the expected timecourse  $\mu$  is constant (i.e.  $I=0$ ), given sufficient statistical metrics. Posterior odds scores are traditionally in log base10, defined as LOD scores. To distinguish a traditional LOD score from  $\log_{10}(O)$ , the authors define this score as the MB-statistic (174). That is,

$$\log_{10}(O) = \text{MB-statistic}$$

Within the full list of 10,227 probe sets eligible for differential expression analysis by the empirical Bayes model, values for the HotellingT2 scores ranged from 5590.9 (probe set: 222912\_at, Arrestin, Beta 1) to 0.02 (31861\_at, immunoglobulin mu binding protein 2) (higher scores are better). The corresponding MB-statistic scores ranged from -0.11 to -7.29. Higher HotellingT2 and MB-statistic scores represent higher confidence that a particular probe set's profile (treated case) is different than its control profile. A full list of probe sets, their HotellingT2 score, and their MB-statistic is provided in **Appendix A**. A MB-statistic cut-off score of -2.0 was chosen to define the subset of differentially expressed probe sets. This cut-off is rather conservative and virtually guarantees the resultant probe-set list as representing true differential expression. Therefore, any probe set with a MB-statistic of -2.0 or greater was included for further analysis. The resulting subset contained 1975 probe sets from probe set 222912\_at "Arrestin Beta 1" with a HotellingT2 score of 5590.9, to 218354\_at

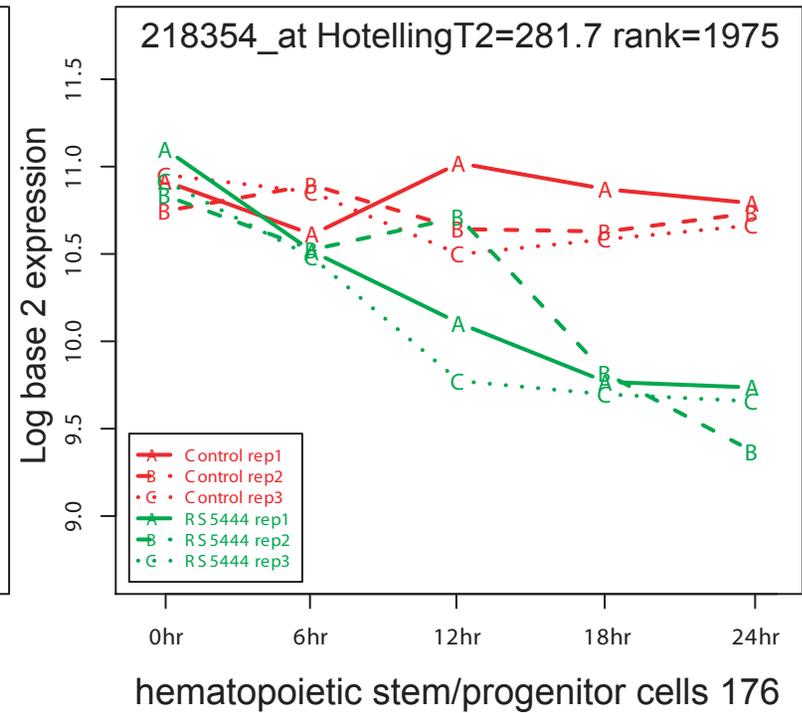
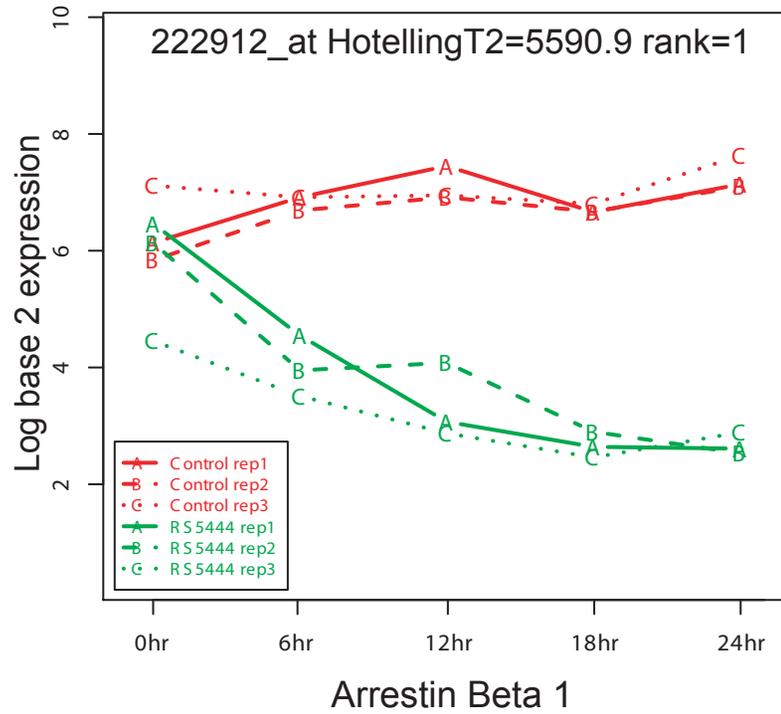
“hematopoietic stem/progenitor cells 176” with a HotellingT2 score of 281.7 [Figure 18].

### **Descriptive statistics of GCRMA summarized expression**

The 1975 differentially expressed probe sets display some interesting general features. Of these probe sets, a majority display down regulation compared to control. If one uses the 24 h time point as a reference point for measuring fold changes between the control and RS5444 treated replicates, 33 probe sets had a positive fold change of +2 or greater and 356 had a negative fold change of -2 or greater. Because the expression data is log base 2, fold change is calculated as 2 to the power of the distance between median control and treated expression. For example, if the median control GCRMA expression of probe set X at time t is 1.0 and the median treatment GCRMA expression of probe set X at time t is 3.0, then the fold change for probe set X at time t is  $2^{(3.0-1.0)} = 2^{2.0} = 4.0$  fold change. If one considers all time points and takes the largest difference of median values at any time point, either negative or positive (defined as the maximum difference of medians or MaxDiff), then 133 probe sets have a positive fold change of +2 or greater and 1807 probe sets have a negative fold change of -2 or greater (more negative), suggesting that the empirical Bayes algorithm models many more genes as down-regulated than up-regulated in the presence of RS5444. As a comparison, the polynomial regression calculated 4,693 probe sets as having a Benjamini-Hochberg adjusted p-value of 0.05 or less. The non-parametric model calculated 6,546 probe sets as having an adjusted p-value of 0.05 or less. Both models calculated about the same percentage of

**Figure 18: *The top and bottom ranked differentially expressed probe sets within the MB-statistic cutoff***

The top ranked and bottom ranked probe sets with an MB-statistic  $> -2.0$  as calculated by the empirical Bayes model. Arrestin Beta 1, with a HotellingT2 score of 5590.9 and a MB-statistic of -0.105 is ranked first of 1975 probe sets. Hematopoietic stem/progenitor cells 176, with a HotellingT2 score of 281.7 and a MB-statistic of -1.998 is ranked last of the 1975 probe sets. Kinetic profiles denote log base 2 GCRMA expression at the times indicated. Red data points denote triplicate samples in the absence of RS5444. Green data points denote triplicate samples in the presence of RS5444.



up-regulated and down-regulated genes as did the empirical Bayes model (data not shown). These results indicate models lacking moderation (see Background) result in what one could assume is moderate to large type-I error.

CEACAM6 probe set “203757\_s\_at”(carcinoembryonic antigen-related cell adhesion molecule 6) which is a known PPAR $\gamma$  target (183), had the greatest positive regulation with a fold change of  $2^{4.8} = 27.9$ . SYNC1 probe set “221276\_s\_at” (syncoilin, intermediate filament 1) had the greatest negative regulation with a fold change of  $2^{-5.24} = -37.8$ . The top twenty positively regulated and negatively regulated probe sets are provided in **Table 8**. The cutoff for **Table 8** was chosen arbitrarily for space considerations.

**Figure 19** summarizes the overall regulation of genes in the presence of RS5444. These data demonstrate that most of the probe sets are down-regulated by PPAR $\gamma$  and is indicative of the processes under the control of RS5444. That is, most functions are down-regulated in the presence of RS5444. Here, the box-plot statistics of probe set’s median distances, at each time point, are shown. The zero hour time point is omitted because the control and treated groups are normalized by the distance at the zero hour time point. Hence, for every probe set at the zero hour time point, this distance is zero. These data support the observation that the empirical Bayes model takes fold change distances across all time points into consideration while calculating the HotellingT2 score. The score is, therefore, a general measure of overall regulation and is positively correlated to the absolute value of the maximum distance between classes [**Figure 20**].

# Table 8: Top Twenty Positively Regulated and Negatively Regulated Probe Sets

## Top 20 Positively Regulated Probe Sets by Maximum Fold Change

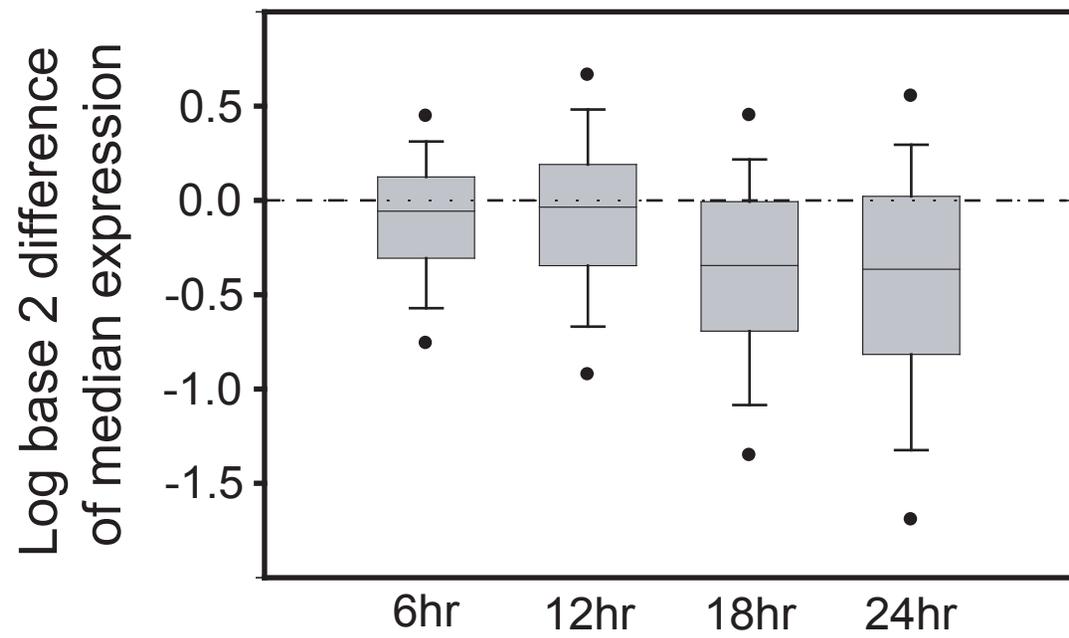
Probe Set	Gene Title	Gene Symbol	HotellingT2 score	MB statistic	Max FC
203757_s_at	carcinoembryonic antigen-related cell adhesion molecule 6	CEACAM6	3958.13	-0.190	4.48
211883_x_at	carcinoembryonic antigen-related cell adhesion molecule 1	CEACAM1	2671.21	-0.320	3.81
209498_at	carcinoembryonic antigen-related cell adhesion molecule 1	CEACAM1	4892.80	-0.135	3.68
212099_at	ras homolog gene family, member B	RHOB	3225.16	-0.253	3.67
224566_at	trophoblast-derived noncoding RNA	TncRNA	1230.80	-0.709	3.63
241940_at	Abhydrolase domain containing 3	ABHD3	1478.04	-0.599	3.22
213816_s_at	met proto-oncogene (hepatocyte growth factor receptor)	MET	1773.12	-0.502	3.06
211657_at	carcinoembryonic antigen-related cell adhesion molecule 6	CEACAM6	2008.34	-0.441	2.93
201650_at	keratin 19	KRT19	1590.98	-0.558	2.91
209122_at	adipose differentiation-related protein	ADFP	2830.39	-0.299	2.88
212531_at	lipocalin 2 (oncogene 24p3)	LCN2	1160.21	-0.747	2.80
214581_x_at	tumor necrosis factor receptor superfamily, member 21	TNFRSF21	2159.26	-0.408	2.77
202672_s_at	activating transcription factor 3	ATF3	1516.05	-0.585	2.74
225239_at	EST	---	1768.71	-0.503	2.71
208370_s_at	Down syndrome critical region gene 1	DSCR1	2256.07	-0.389	2.69
209365_s_at	extracellular matrix protein 1	ECM1	1792.35	-0.496	2.62
212806_at	KIAA0367	KIAA0367	1586.62	-0.560	2.58
223541_at	hyaluronan synthase 3	HAS3	1897.84	-0.468	2.57
217173_s_at	low density lipoprotein receptor	LDLR	1140.01	-0.758	2.51
202842_s_at	DnaJ (Hsp40) homolog, subfamily B, member 9	DNAJB9	997.23	-0.849	2.44

## Top 20 Negatively Regulated Probe Sets by Maximum Fold Change

Probe Set	Gene Title	Gene Symbol	HotellingT2 score	MB statistic	Max FC
212570_at	KIAA0830 protein	KIAA0830	1086.05	-0.790	-3.93
234987_at	Chromosome 20 open reading frame 118	C20orf118	3697.08	-0.210	-3.93
225665_at	sterile alpha motif and leucine zipper containing kinase AZK	ZAK	717.36	-1.100	-3.96
215029_at	EST	---	2621.56	-0.327	-3.99
228250_at	KIAA1961 gene	KIAA1961	5132.36	-0.124	-4.00
225773_at	KIAA1972 protein	KIAA1972	2709.89	-0.315	-4.00
230146_s_at	frequenin homolog (Drosophila)	FREQ	2097.36	-0.421	-4.03
213357_at	general transcription factor IIH, polypeptide 5	GTF2H5	2108.35	-0.419	-4.09
205471_s_at	dachshund homolog 1 (Drosophila)	DACH1	3840.05	-0.199	-4.16
212249_at	phosphoinositide-3-kinase, regulatory subunit 1 (p85 alpha)	PIK3R1	1619.61	-0.549	-4.26
218692_at	hypothetical protein FLJ20366	FLJ20366	678.65	-1.146	-4.33
201470_at	glutathione S-transferase omega 1	GSTO1	3462.55	-0.230	-4.34
212131_at	family with sequence similarity 61, member A	FAM61A	1926.76	-0.461	-4.39
203255_at	F-box protein 11	FBXO11	407.42	-1.614	-4.39
223984_s_at	nucleoporin like 1	NUPL1	2822.08	-0.300	-4.42
222912_at	arrestin, beta 1	ARRB1	5590.89	-0.105	-4.53
1555772_a_at	cell division cycle 25A	CDC25A	4091.39	-0.181	-4.61
212386_at	Transcription factor 4	TCF4	2164.22	-0.407	-4.72
205472_s_at	dachshund homolog 1 (Drosophila)	DACH1	2365.74	-0.369	-5.01
221276_s_at	syncollin, intermediate filament 1	SYNC1	4188.57	-0.175	-5.24

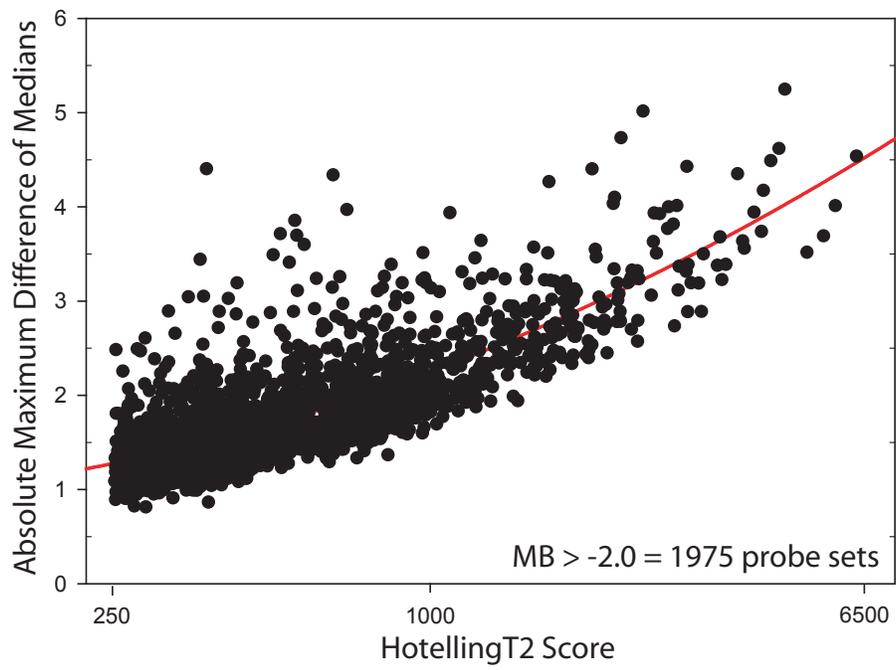
**Figure 19: *A majority of probe sets are down-regulated by PPAR $\gamma$***

Box-and-whisker plot of median distances between control and treated groups of the differentially expressed probe sets. All probe set median distances are normalized to the distance of the zero hour time point. Hence, data from the zero hour time point is omitted. Box plot displays the five-number summary. That is, boxes represent median and upper/lower quartiles. Whiskers represent smallest and largest observation. In addition, dots represent median of outliers.



**Figure 20: *HotellingT2 score is positively correlated to regulation***

Semi-log plot of HotellingT2 scores versus the absolute value of the maximum difference between median expression values of control versus treated condition at any time point for each of the 1975 differentially expressed probe sets.



### **Time course kinetic classes**

The 1975 differentially expressed probe sets can be grouped into five classes of expression kinetics. **Figure 21A** depicts the class of down-regulated genes, achieving steady-state with probe set 222912\_at “Arrestin Beta 1.” This probe set achieved steady-state transcript expression within 24 h after addition of RS5444. **Figure 21B** depicts the class of up-regulated genes, achieving steady-state with the probe set 214581\_x\_at “tumor necrosis factor receptor superfamily, member 21.” The next two classes represent genes which display linear kinetics. These probe sets did not achieve steady-state regulation. **Figure 22A** depicts the up-regulated probe set 202199\_at “Ras homolog gene family, member B” from this class and **Figure 22B** depicts the down-regulated probe set 205471\_s\_at “Dachshund homolog 1” from this class. **Figure 23** depicts the class of probe-sets where regulation can not be explained. The example given is probe set 1567213\_at “Pinin, desosome associated protein” which displays regulation in the control group, and not in the treated set.

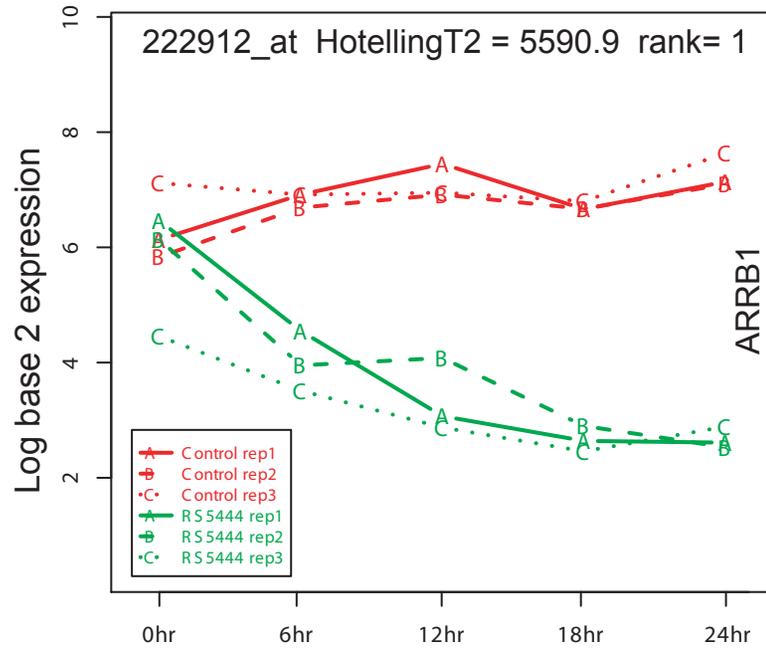
### **Ontology analysis of gene expression**

The program ermineJ was used to analyze gene ontology (GO) of the 1975 differentially expressed probe sets identified by the empirical Bayes HotellingT2 model as described in Materials and Methods. Of the 6127 GO classes considered, 112 GO classes were found to have a p-value of 0.05 or less by over-representation analysis. A full list of GO classes can be found in **Appendix B**. Interestingly, some of the highest scoring GO classes involved many functional aspects of proliferation and replication

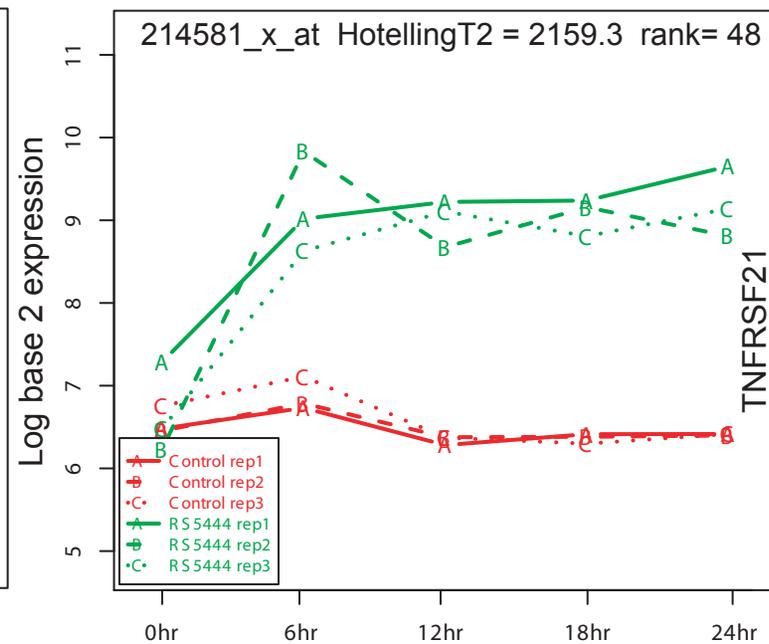
**Figure 21: *Time course kinetic classes representing steady-state regulation***

Probe sets representing a general observation of steady-state regulation. **A)** Representative probe set 222912\_at “Arrestin Beta 1 a.k.a. ARRB1” displaying down-regulated kinetics achieving steady-state within 24 h after RS5444 treatment. **B)** Representative probe set 214581\_x\_at “tumor necrosis factor receptor superfamily, member 21 a.k.a. TNFRSF21” displaying up-regulated kinetics achieving steady-state within 24 h after RS5444 treatment.

A. Down-regulated probeset achieving steady-state



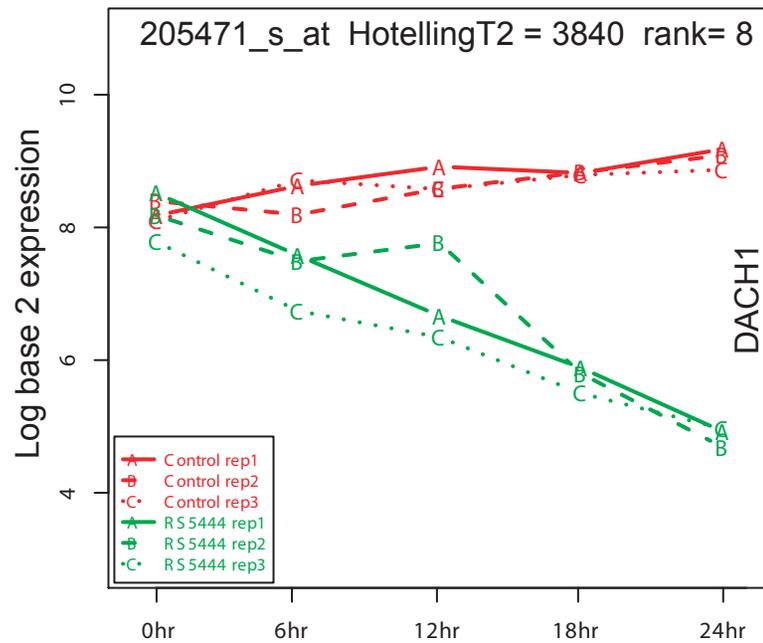
B. Up-regulated probeset achieving steady-state



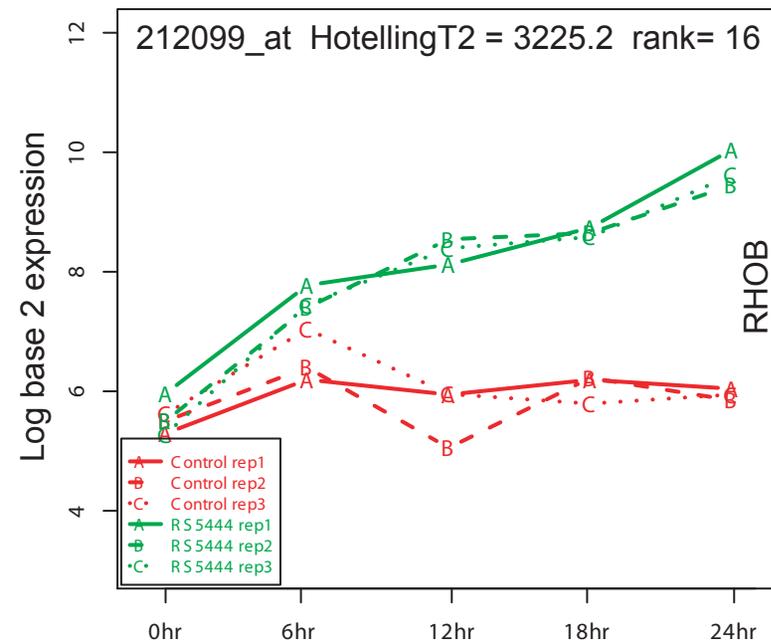
**Figure 22: *Time course kinetic classes representing linear, non- steady-state regulation***

Probe sets which represent a general observation of non- steady-state regulation. **A)** Representative probe set 202199\_at “Ras homolog gene family, member B a.k.a. RHOB” displaying down-regulated kinetics which did not achieve steady-state within 24 h after RS5444 treatment. **B)** Representative probe set 205471\_s\_at “Dachshund homolog 1 a.k.a DACH1” displaying up-regulated kinetics which did not achieve steady-state within 24 h after RS5444 treatment.

A. Down-regulated probe-set not achieving steady-state

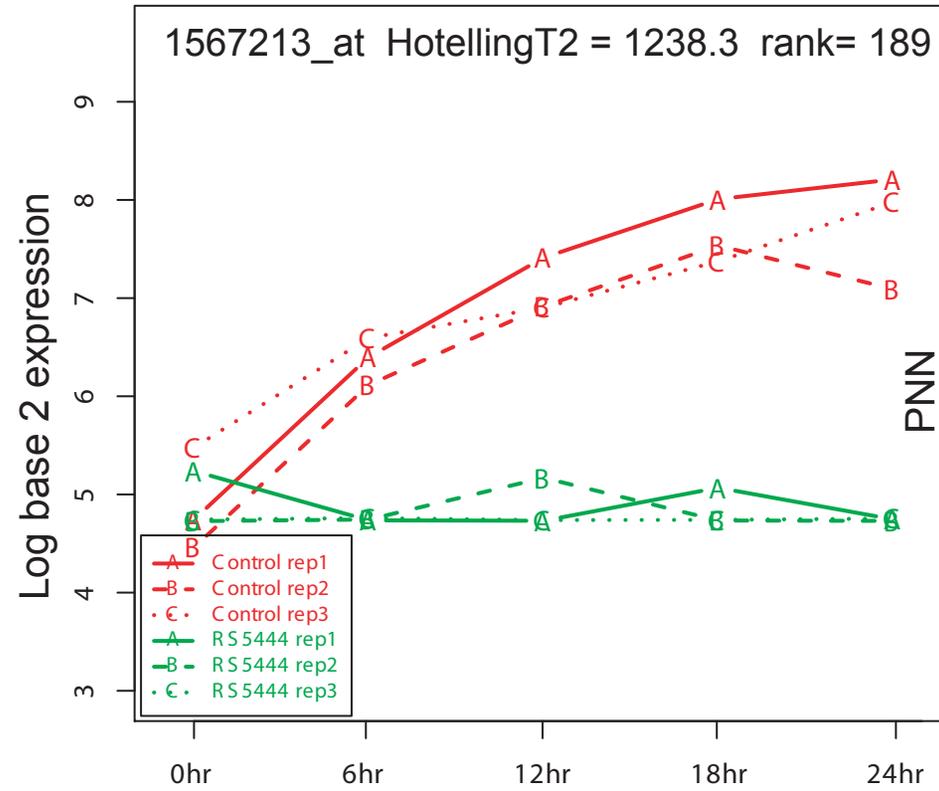


B. Up-regulated probe-set not achieving steady-state



**Figure 23: *Time course kinetic classes representing regulation which can not be explained***

Representative probe set 1567213\_at “Pinin, desosome associated protein a.k.a PNN” displaying kinetics where the control set is regulated in the absence of RS5444

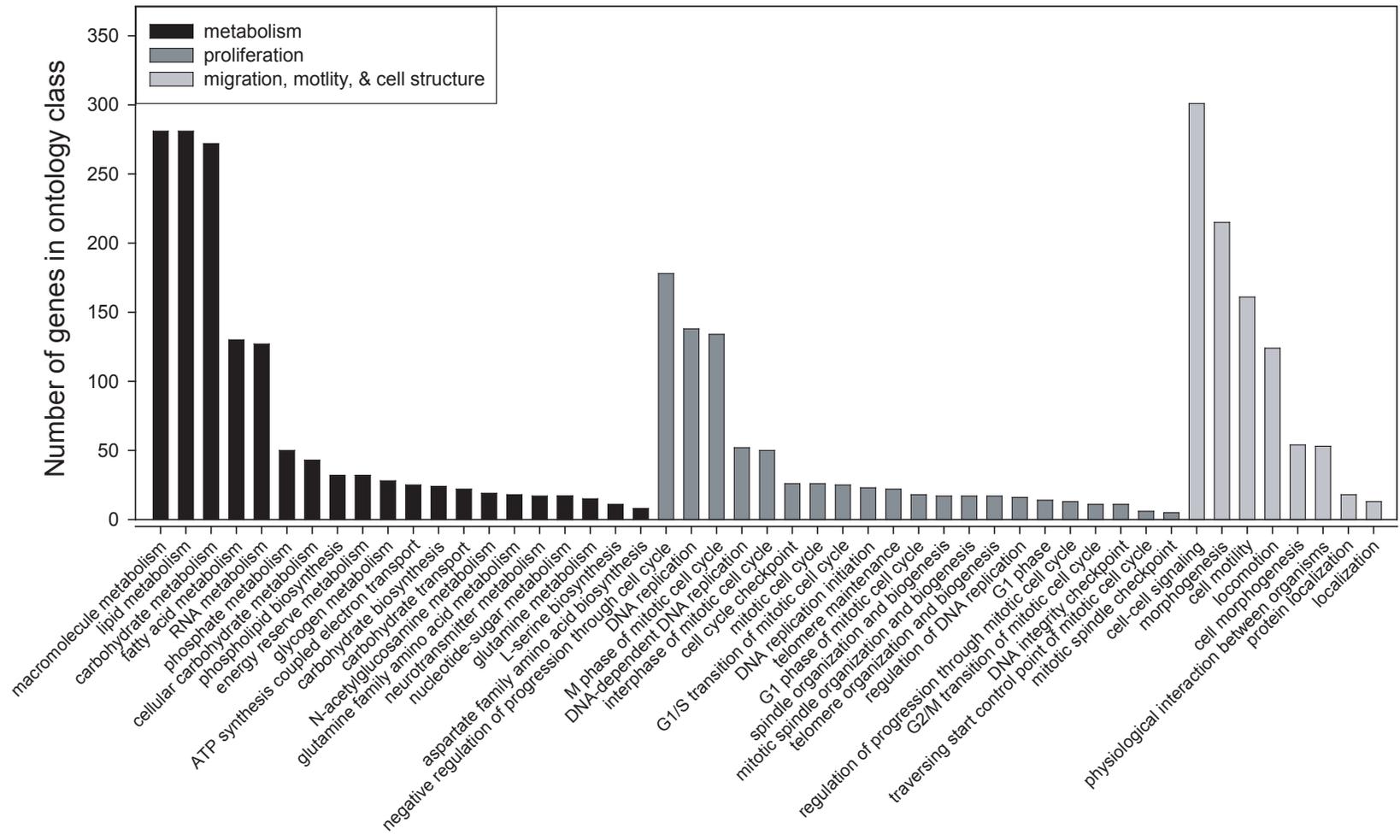


functions. These data are consistent with our observation that PPAR $\gamma$  controls proliferation in colorectal cancer cells [Figure 9] and [Figure 10]. Gene ontology analysis also revealed that many classes involving metabolism functions were also regulated. These data are consistent with known metabolic control functions of PPAR $\gamma$  in a wide variety of cell types (187-192). Finally we found several GO classes significantly regulated which are involved in migration, motility, and cell structure. Each of these functions is important in invasion. This observation is consistent with our observation that PPAR $\gamma$  controls invasion in colorectal cancer cells [Figure 12]. They are also consistent with data published in our laboratory involving motility in normal intestinal epithelial cells (79, 109). Specific GO terms within these superclasses are shown in Figure 24 along with the number of differentially regulated genes within each GO term. It should be noted here that many GO terms have subtle differences in definition and as a consequence, a specific gene may be represented more than once in this kind of analysis.

It is informative to look at GO terms representing significantly regulated functions within each of these superclasses. Again, metabolism, proliferation, and migration/motility were all functions we expected to appear in this analysis. In this respect the gene ontology analysis was used as a kind of reality-check to ensure that our 1975 probe sets represented regulated biological functions that matched what was observed during our biological characterization experiments. For example, as would be expected, lipid metabolism [Figure 25] which is a prototypical function of PPAR $\gamma$  shows

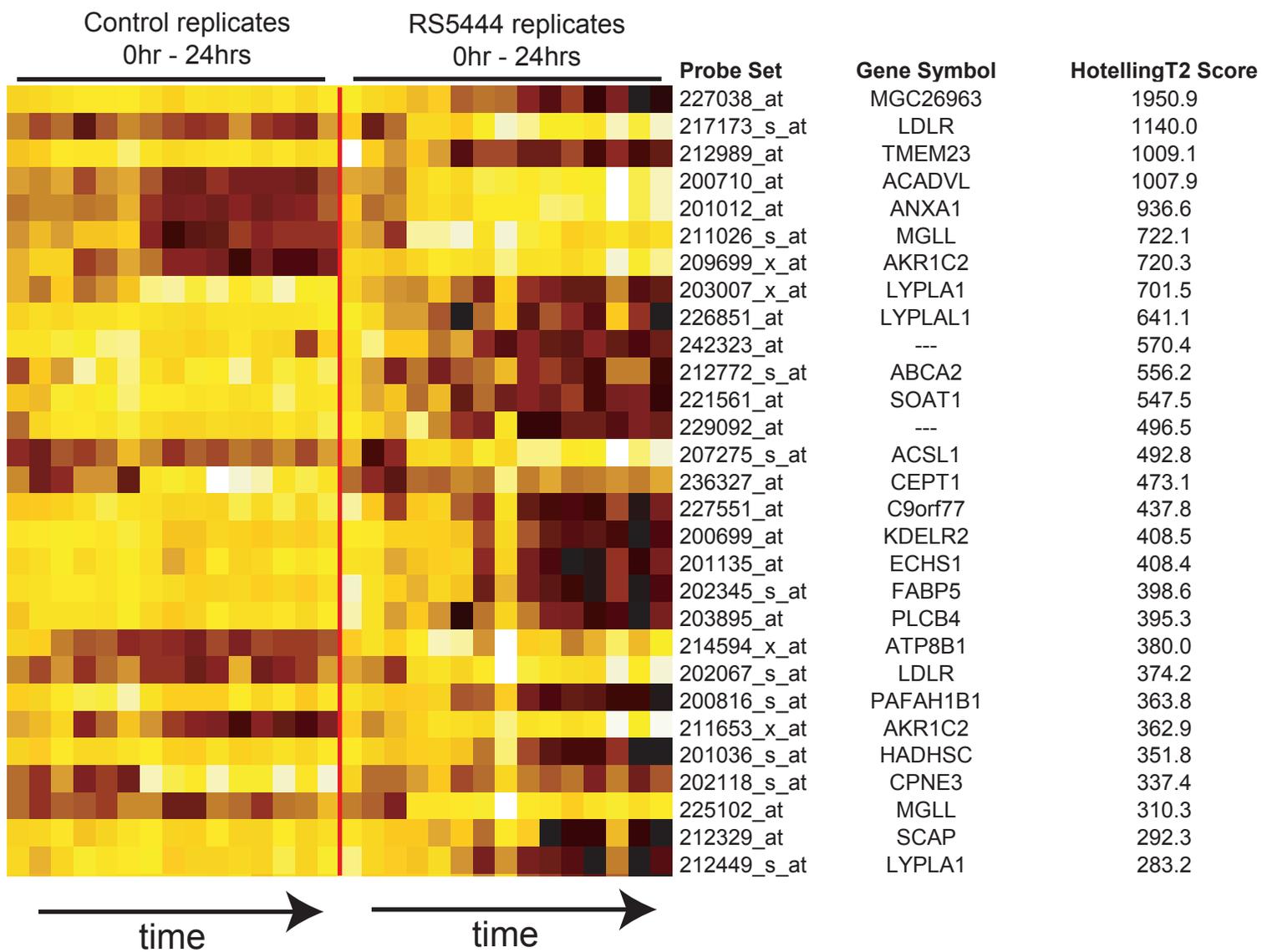
**Figure 24: *Gene ontology classes matching known functions of PPAR $\gamma$***

The list of GO terms calculated by the program ermineJ grouped together into three superfamilies of biological function expected as a consequence of treatment with RS5444. Each GO term meets or exceeds an ermineJ calculated p-value of  $< 0.05$ . The first superfamily, metabolism, consists of twenty GO terms here force ranked by the number of genes within each term. The second superfamily, proliferation, consists of twenty-one GO terms. The third superfamily, which is a composite of several GO terms making up biological functions involved in invasiveness, consists of eight GO terms



**Figure 25: *Lipid metabolism: prototypical GO term subclass of the metabolism superfamily regulated by PPAR $\gamma$***

A heat map of regulated probe sets, force ranked by their HotellingT2 score, representing genes that are involved in aspects of lipid metabolism, which is grouped under the superfamily of metabolism. The data are represented using dark-body coloring normalized around mean expression where darker colors represent lower expression values, and lighter colors represent higher expression values. For example, the probe set 227038\_at “MGC26963” is down-regulated over 24 hours in the presence of RS5444.

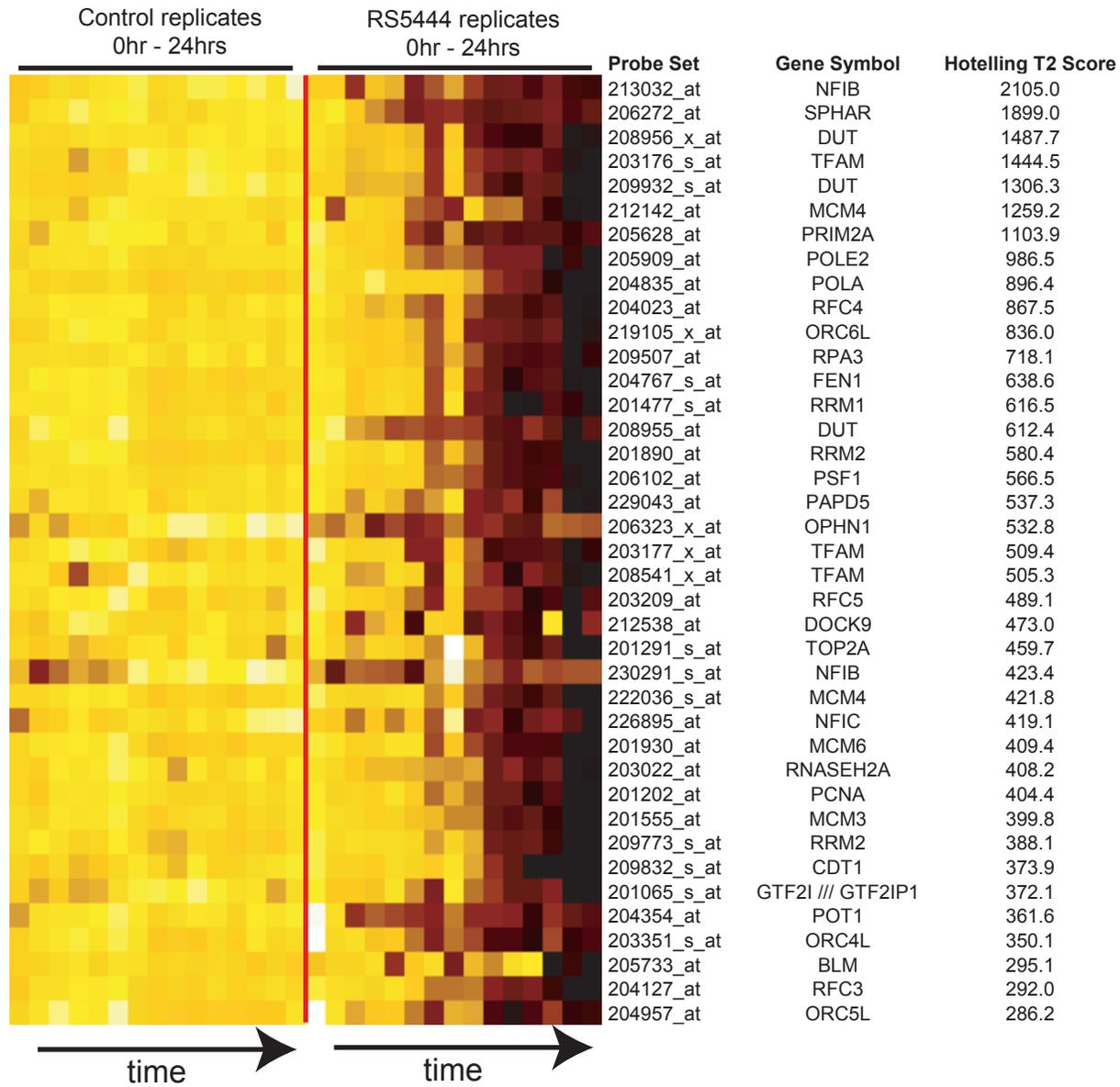


both strongly positive and strongly negative regulated probe sets representing known PPAR $\gamma$  targets involved in metabolism. Another example is DNA replication [Figure 26]. As can be seen, every significant probe set within this class is down-regulated, which confirms the phenotypic effects of PPAR $\gamma$  on proliferation. Finally, much like lipid metabolism, probe sets involved in migration, motility and cell-structure show both positively and negatively regulated probe sets which are consistent with the observation that PPAR $\gamma$  appears to regulate aspects involved in invasion, but not motility in colorectal cancer cells. The GO term subclass morphogenesis [Figure 27] illustrates this point.

Much to our surprise, gene ontology analysis provided two classes which have no previous association with PPAR $\gamma$  in colorectal cancer cells: angiogenesis and calcium-mediated signaling. The cohort of genes which regulate angiogenesis surprised us because it is not a known property of tumor cells [Figure 28A]. Angiogenesis is a property of endothelial cells. However, if one considers the processes that are involved in angiogenesis, they include changes in adhesion, migration, motility, and invasion—all of which are high scoring functions via ermineJ analysis. Calcium-mediated signaling was not expected either [Figure 28B]. To our knowledge, no one has shown a relationship between PPAR $\gamma$  and calcium-mediated signaling, although there is a considerable body of evidence that links dietary calcium intake to suppression of colon cancer (193-196).

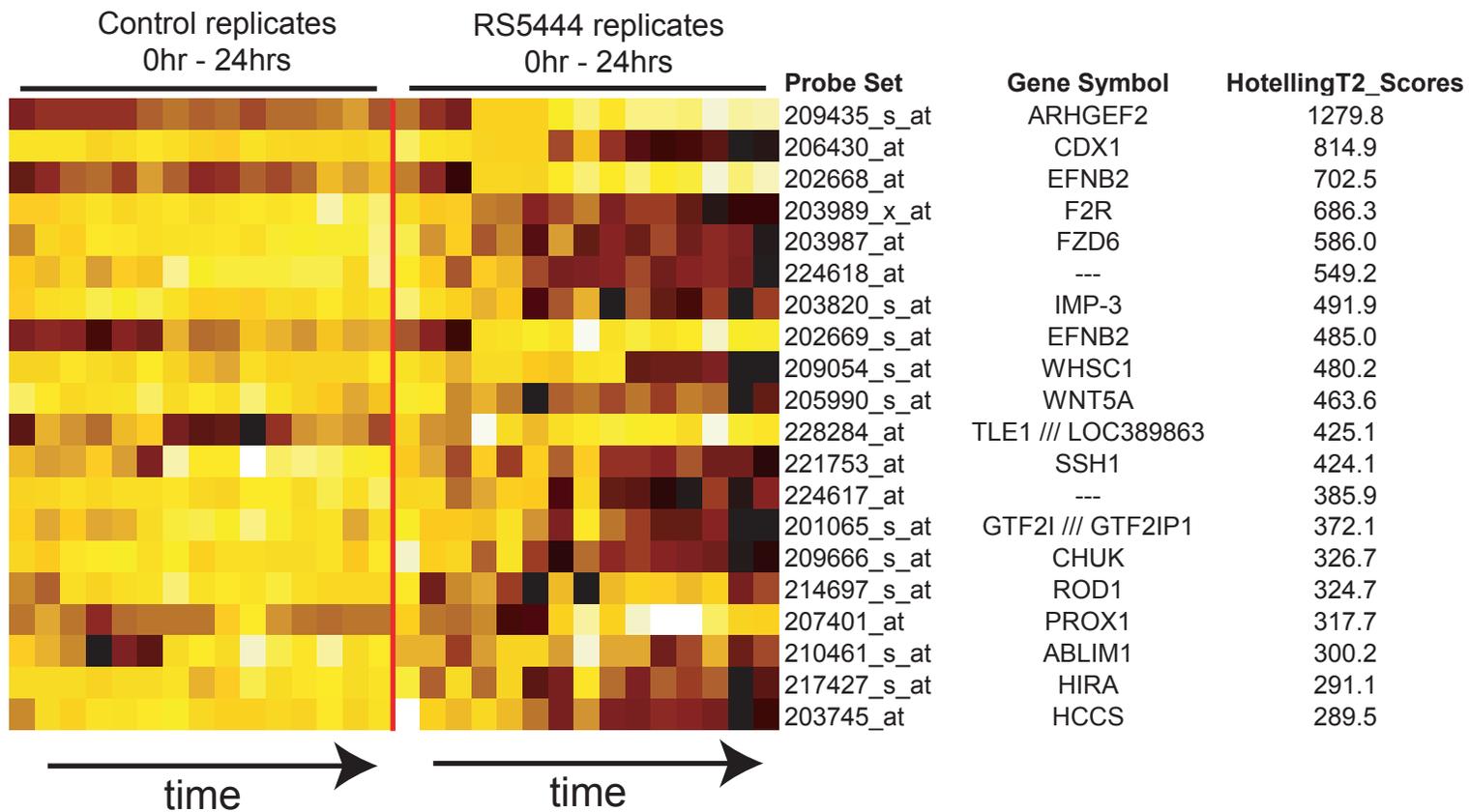
**Figure 26: *DNA replication: prototypical GO term subclass of proliferation regulated by PPAR $\gamma$***

A heat map of regulated probe sets involved in aspects of DNA replication, which is a grouped under the superfamily of proliferation. Probe sets exceed the MB-statistic cutoff and are force ranked by HotellingT2 score. All probe sets in this GO term subclass are strongly down-regulated in the presence of RS5444 over 24 hours.



**Figure 27: *Morphogenesis: prototypical GO term subclass of invasion regulated by PPAR $\gamma$***

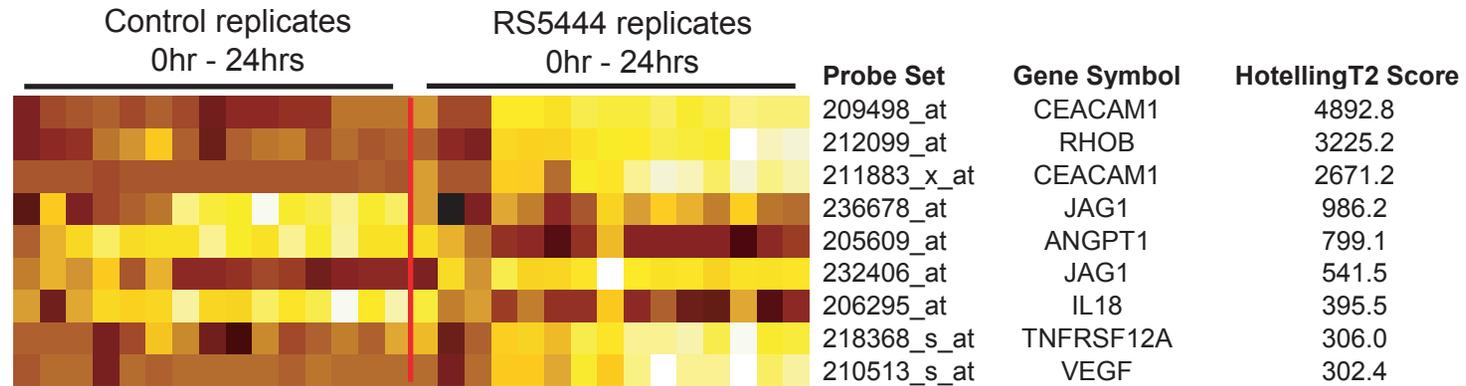
A heat map of regulated probe sets involved in aspects of morphogenesis, which is grouped under a collection of functional classes involved in invasiveness. Probe sets exceed the MB-statistic cutoff, and are forced ranked by HotellingT2 score.



**Figure 28:** *Angiogenesis and Calcium-mediated Signaling: ontology classes not expected to be regulated by PPAR $\gamma$*

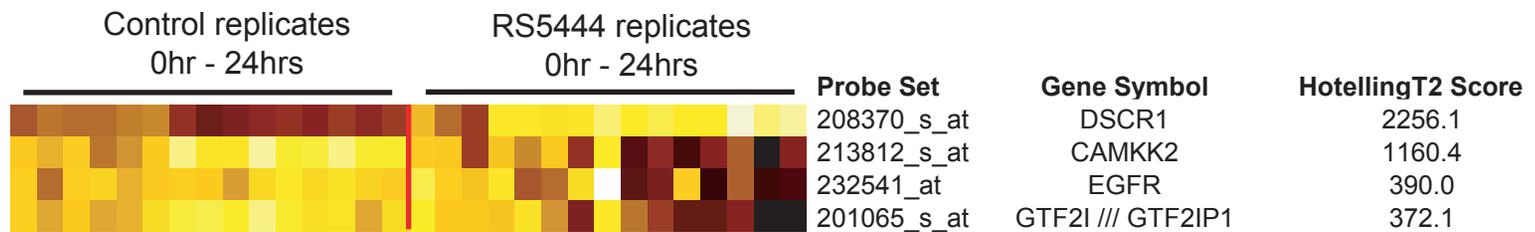
**A)** A heat map of the cohort of probe sets making up the angiogenesis GO term. Probe sets exceed the MB-statistic cutoff, and are forced ranked by HotellingT2 score. **B)** A heat map of the cohort of probe sets making up the calcium-mediated signaling GO term. Probe sets exceed the MB-statistic cutoff, and are forced ranked by HotellingT2 score.

A.



Angiogenesis GO term sub-class

B.



Calcium-mediated signaling GO term sub-class

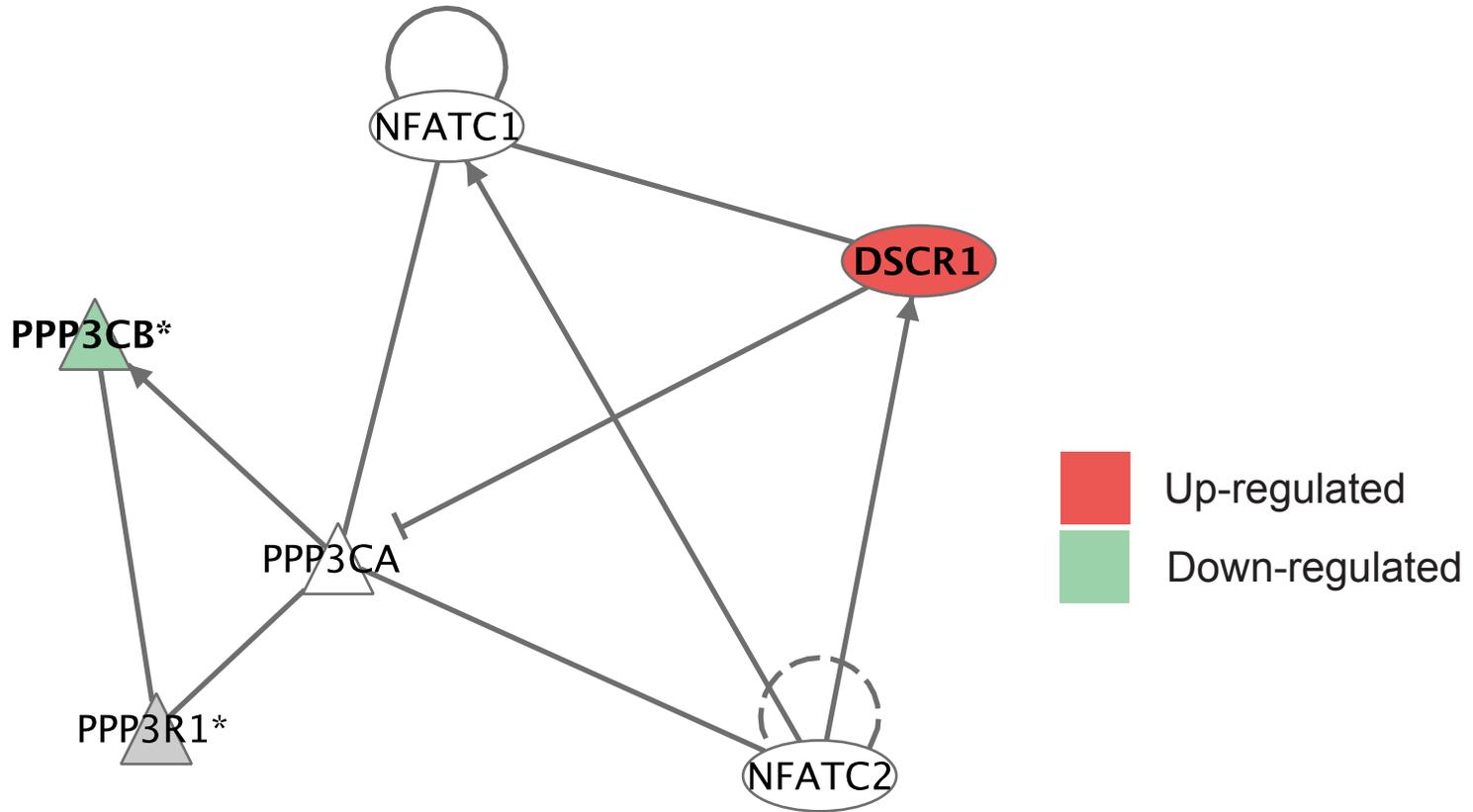
### **Ingenuity Pathways Analysis infers a DSCR1-calcineurin-NFATc signaling axis**

The set of 10277 probe sets with MAS5 change calls of “P” or “M” across all conditions were uploaded into Ingenuity Pathways Knowledge Base (IPKB). Accompanying the probe sets were their associated HotellingT2 scores and the maximum difference of medians distance (explained in the descriptive statistics section above) as figures of merit for significance calculation. Using Ingenuity Pathways Analysis (IPA) functions, probe sets from the angiogenesis and calcium-mediated signaling GO term sets were manipulated within IPA to find possible connections to the observations we made during our biological characterization of MOSER S cells. **Figure 29** summarizes our initial findings. IPA inferred a novel relationship between the calcineurin-inhibitor DSCR1, and two other genes which did not show up in any of our other bioinformatic analyses: the transcription factor Nuclear factor of activated T-cells (cytoplasmic component) (NFATC1 and C2) and its regulator calcineurin (PPP3CA and B), which is a serine-threonine protein phosphatase. This schematic models DSCR1 as calcineurin’s inhibitor.

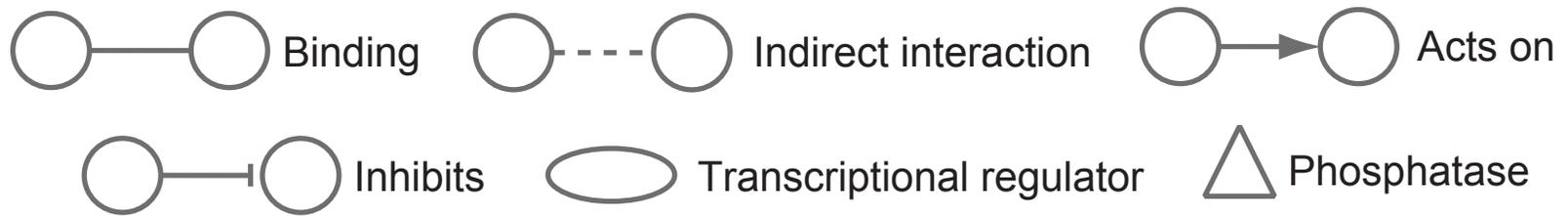
Reviewing the literature showed that these genes were members of the calcium-mediated signaling canonical pathway. Overlaying known members of this pathway stored in Ingenuity’s knowledge base with HotellingT2 scores reiterates this point [**Figure 30**]. Notice that all members of this canonical pathway are down-regulated (indicated by green) except for DSCR1, which again is an inhibitor and is induced (red).

**Figure 29: *Ingenuity Pathways Analysis Inferred DSCR1-Calciuneurin-NFATc pathway mediated by PPAR $\gamma$***

Ingenuity Pathways Analysis output of genes involved in calcium-mediated signaling. Colored genes represent maximum fold change overlay as described above. Following heatmap coloring convention, red represents up-regulation and green represents down-regulation. Genes which were part of the initial 10277 input probe sets, but which did not make the MB-statistic cutoff, are displayed in grey. Genes which were not part of the 10277 input probe sets, but were determined important by Ingenuity Pathways Knowledge Base for completion of the model, are shown in white. The legend of the various connections between genes is provided.

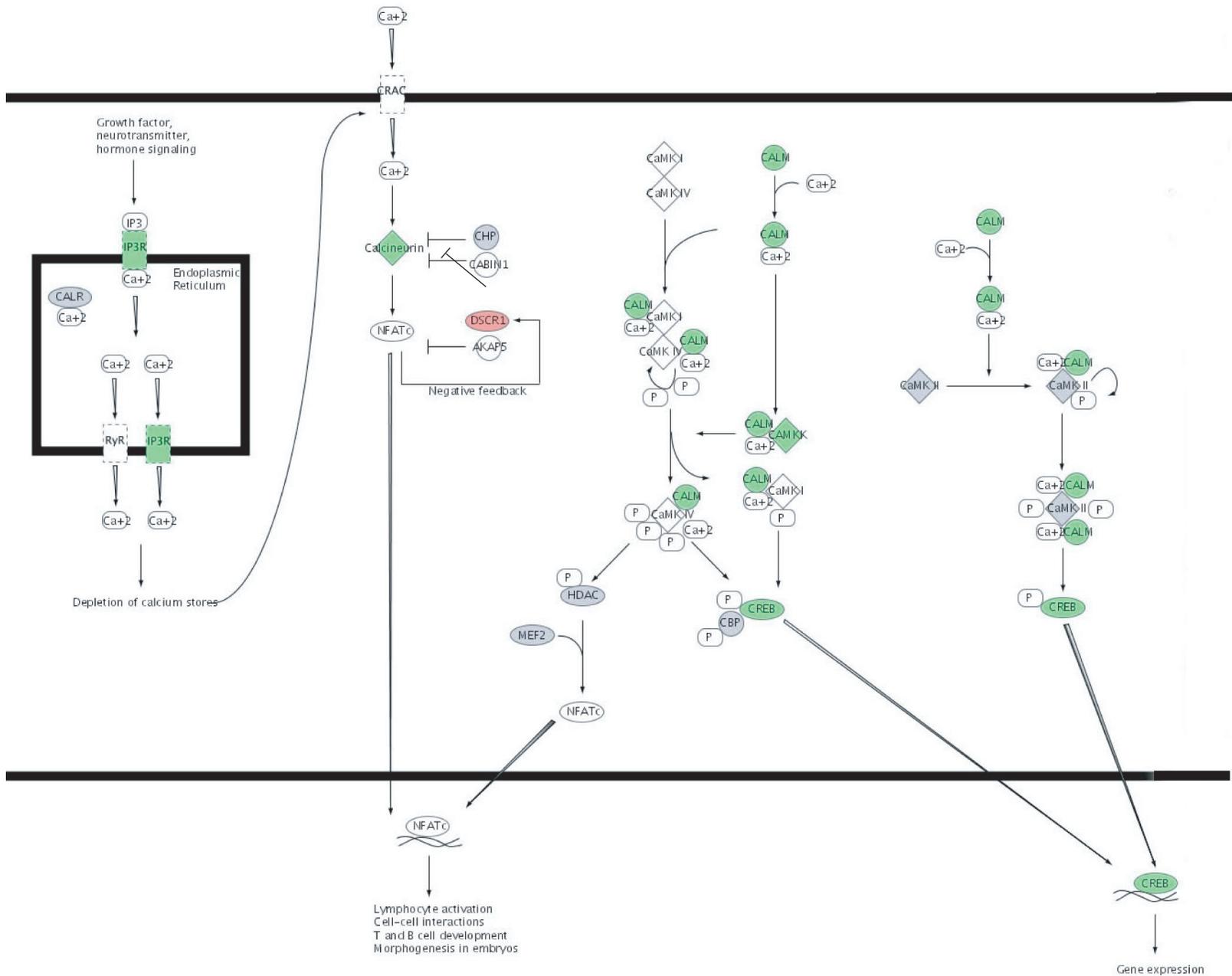


Up-regulated  
 Down-regulated



**Figure 30: *Ingenuity Pathways Knowledge Base canonical pathway of calcium-mediated signaling with PPAR $\gamma$  regulation overlay***

The calcium-mediated signaling canonical pathway stored in the Ingenuity Pathways Knowledge Base overlaid with maximum fold change. All colored genes represent probe sets making the MB-statistic cutoff. All grey genes represent probe sets which were part of the initial 10277 input probe sets, but did not make the MB-statistic cutoff. All white genes were not part of the initial 10277 input probe sets.

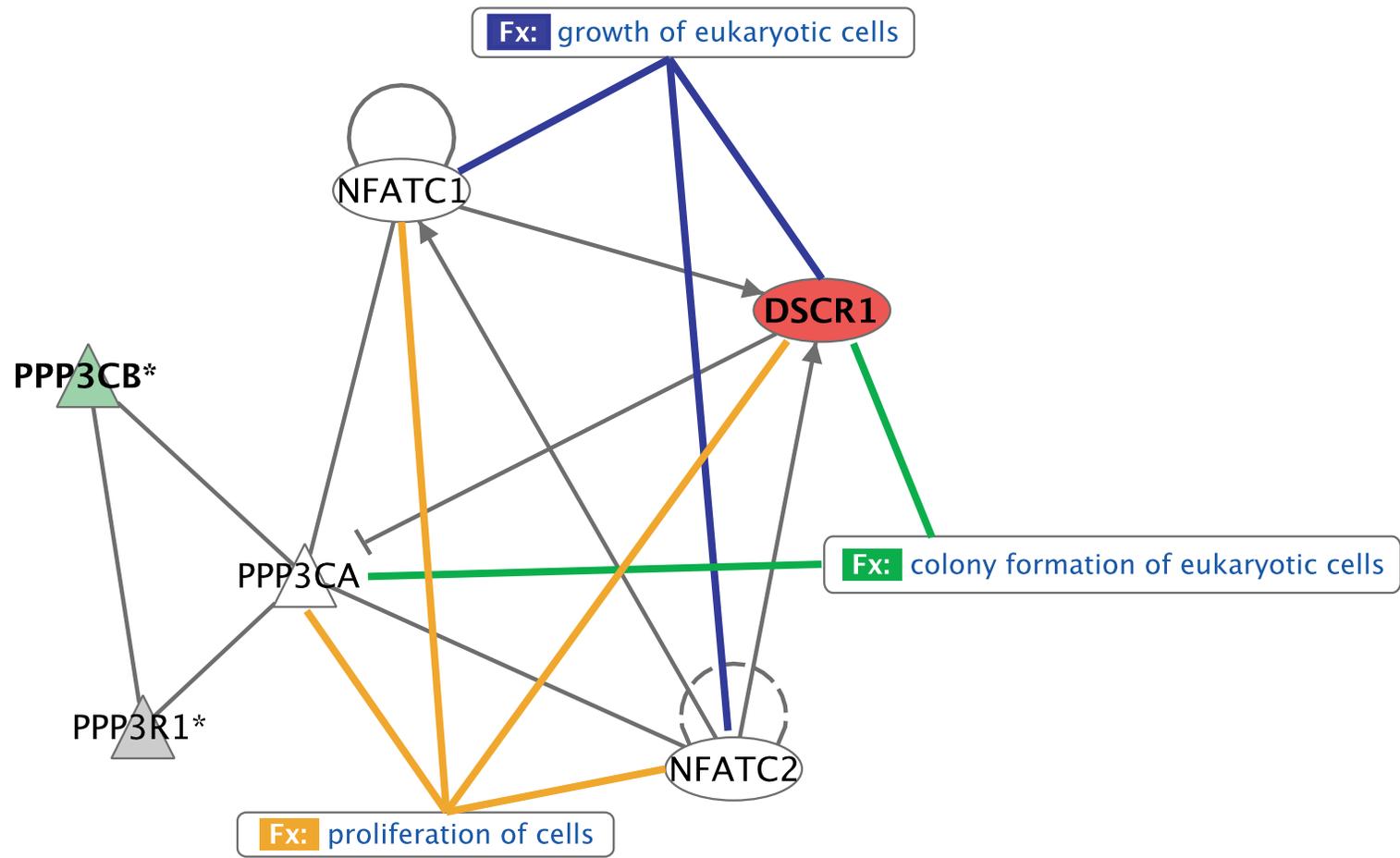


Next, the gene candidates in **Figure 29** were overlaid with functions representing our initial biological characterization, i.e. functions involved in proliferation and functions involved in invasion. **Figure 31** summarizes the findings for gene candidates with known functions involved in proliferation. The matches include functions involving growth of eukaryotic cells (139, 197-199), colony formation of eukaryotic cells (134, 200), and proliferation in cells (139, 201-203). **Figure 32** summarizes the finding for gene candidates with known functions involved in invasion. The matches include invasion of colon cancer cell lines (142), tubulation of endothelial cells (139), and area of myofiber (198).

Our genomic analyses predict that PPAR $\gamma$  may regulate calcium-mediated signaling involving calcineurin, NFATc, and DSCR1. A review of the known functions of NFATc in carcinogenesis suggests that this pathway may be involved in regulating PPAR $\gamma$  induced suppression of proliferation and invasion of MOSER S colorectal cancer cells. The DSCR1 gene (also known as Adapt78 or calsupressin) consists of four different isoforms as a result of alternative first exon promoter usage. However, only isoforms 1, 2, and 4 have been identified (132). We designed PCR primers to determine which of the three isoforms are expressed in MOSER S cells. As shown in **Figure 33**, differential PCR primers were designed to distinguish between isoforms 1 and 2/4. We were unable to design isoform 2 specific primers because the nucleotide sequence of isoform 2 is included in both isoforms 1 and 4. Instead, we designed a primer set that can recognize all three isoforms (1, 2, and 4). Our results show that while isoform 1 is not

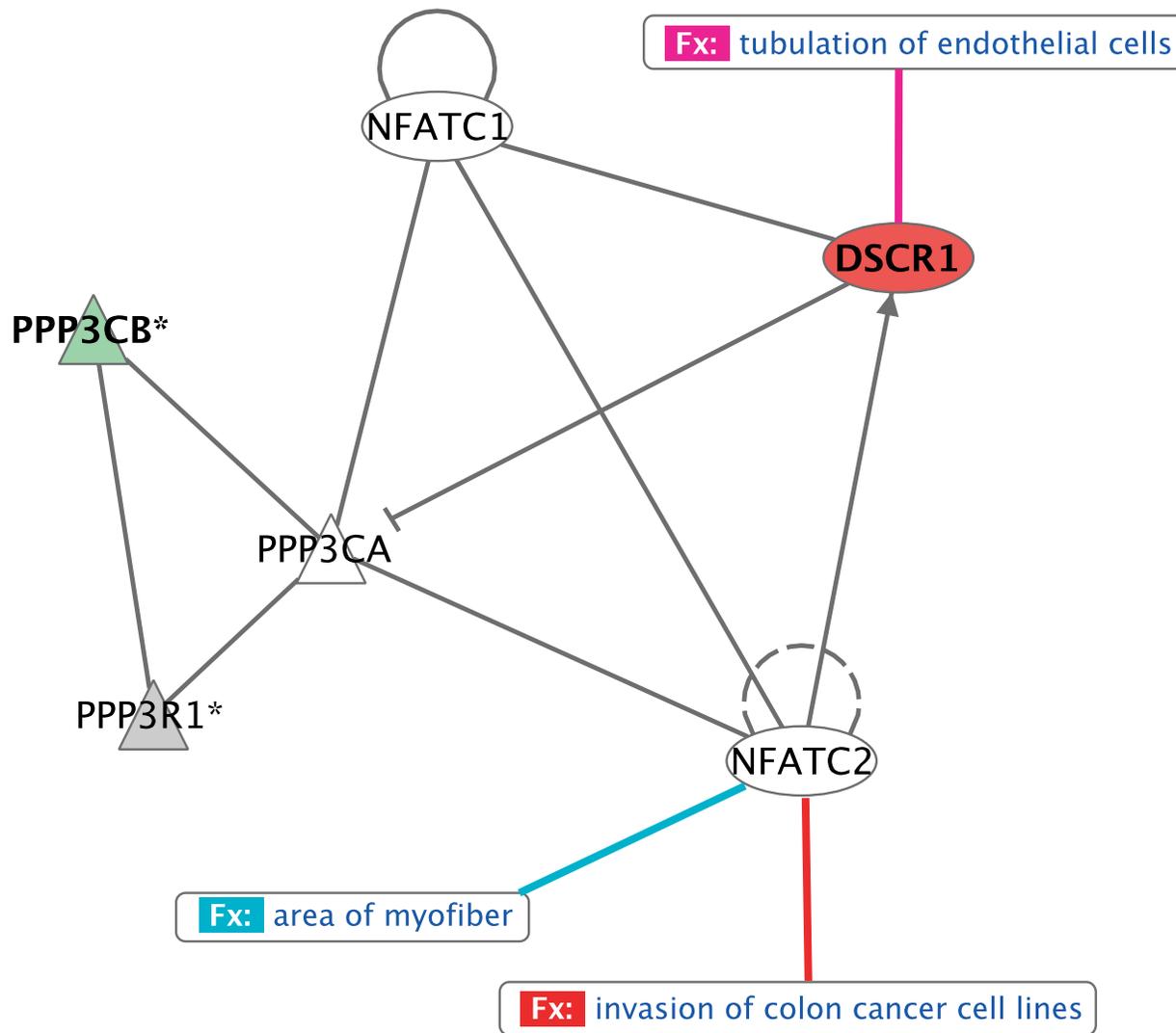
**Figure 31: *Functions involved in proliferation overlaid on DSCR1-Calciuneurin-NFATc inferred model***

The inferred calcium-mediated signaling axis of genes [Figure 29], overlaid with biological functions of proliferation, represented in the literature. Dark Blue: growth in eukaryotic cells. Green: colony formation in eukaryotic cells. Orange: proliferation of cells.



**Figure 32: *Functions involved in invasiveness overlaid on DSCR1-Calciuneurin-NFATc inferred model***

The inferred calcium-mediated signaling axis of genes [Figure 29], overlaid with biological functions of invasiveness, represented in the literature. Magenta: tabulation of endothelial cells. Cyan: area of myofiber. Red: invasion of colorectal cancer cells.



**Figure 33: *DSCR1* isoform 1 is expressed in MOSER S cells**

**A)** The published cDNA sequence of *DSCR1* isoforms 1 (NM\_004414), 2 (NM\_203417), and 4 (NM\_203418). The forward and reverse PCR primer sets to detect the different isoforms in MOSER S cells are highlighted (bolded/underlined). *DSCR1* isoform 1 has a unique 5' sequence illustrated in green. *DSCR1* isoform 4 has the unique 5' sequence illustrated in magenta. *DSCR1* isoform 2's coding sequence is present in both *DSCR1* isoforms 1 and 4 (red) and therefore does not present a unique sequence to detect by PCR. **B)** Total RNA from MOSER S treated with or without RS5444 for 24 h was reverse transcribed into cDNA. A PCR reaction was performed using primer set #1 (detects *DSCR1* isoform 1 alone), primer set #2 (detects *DSCR1* isoform 4 alone), or primer set #3 (detects all three *DSCR1* isoforms)



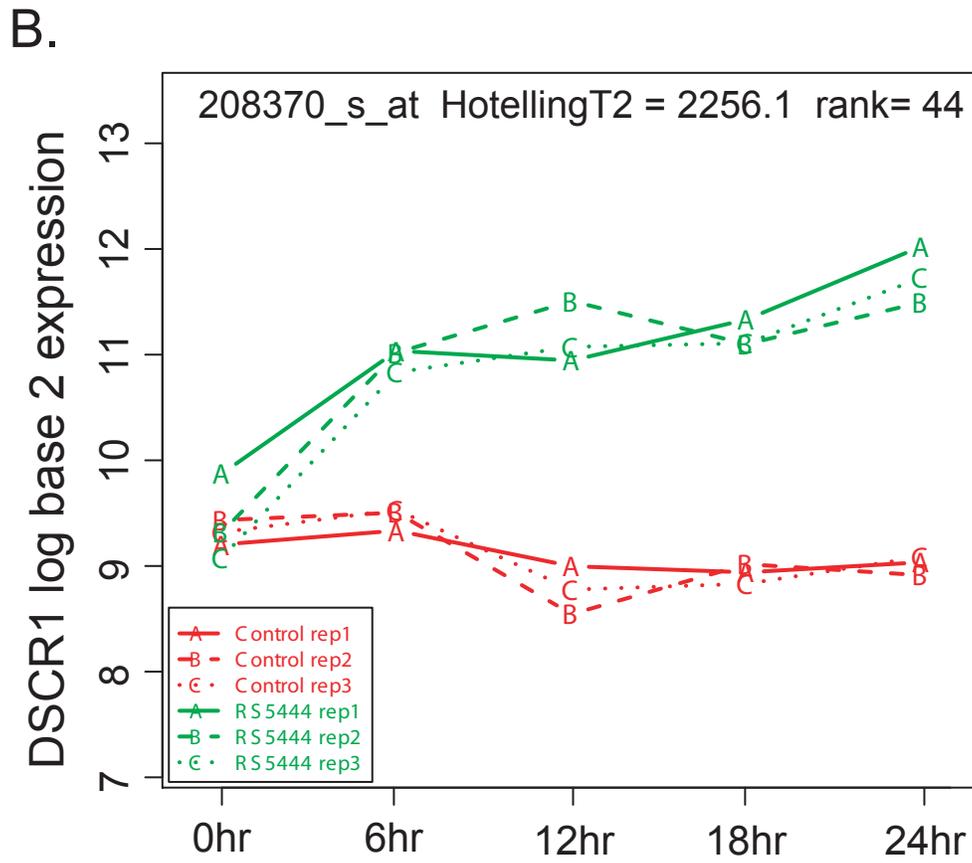
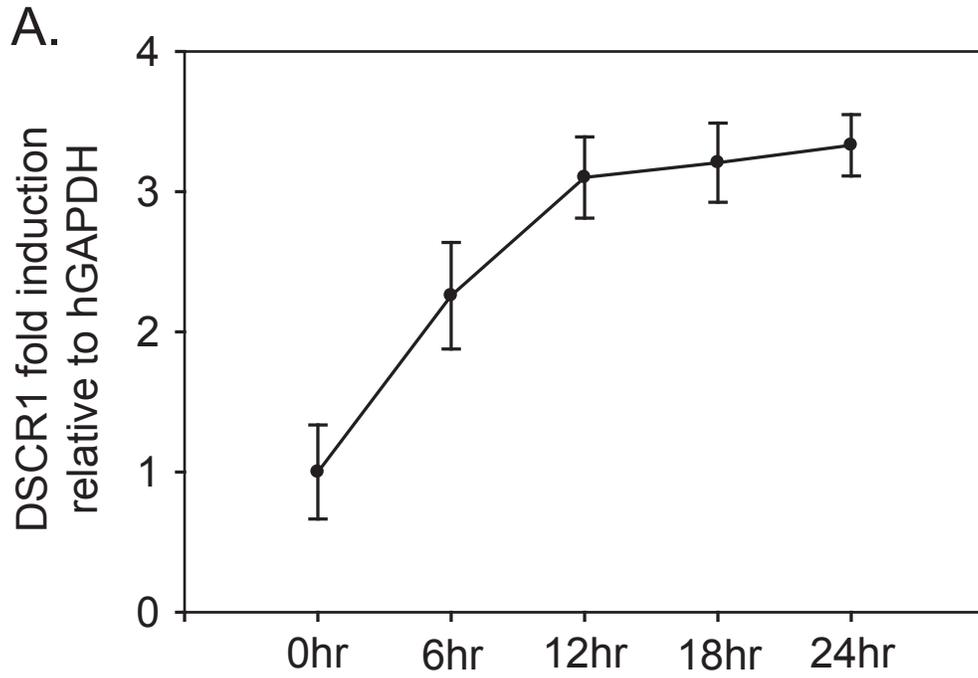
detectable, isoform 4 (or isoforms 2 and 4) is present in MOSER S cells. Although to our knowledge no one has previously demonstrated that DSCR1 isoform 4 is expressed in the colon, others have shown that isoform 4 is expressed in a relatively wide array of tissue types (132).

We next used qPCR to verify the HotellingT2 kinetic model of DSCR1 mRNA induction following PPAR $\gamma$  activation. The HotellingT2 model predicts that RS5444 induces DSCR1 mRNA expression by as much as 3-fold. Consistent with this prediction, DRCS1 mRNA measured by qPCR showed a 3-fold induction upon RS5444 treatment [Figure 34]. Moreover, the HotellingT2 kinetic model predicts that DSCR1 mRNA is rapidly induced by RS5444 and reaches steady state 12 h after RS5444 treatment. Our qPCR data showed a similar temporal expression pattern of DSCR1, confirming the kinetics predicted by the Hotelling T2 model.

The induction of DSCR1 by thiazolidinediones is dose-dependent. The abundance of DSCR1 mRNA in MOSER S cells treated with increasing concentration of RS5444 or rosiglitazone was measured by qPCR [Figure 35]. The concentrations of RS5444 and rosiglitazone required to achieve one-half of the maximum induction of DSCR1 mRNA are comparable to those required for PDK4 mRNA [Figure 6C] and [Table 5]. The results do not perfectly correlate with PDK4 and this may be due to the endogenous concentrations of transcript. PDK4 is used in our laboratory as a PPAR $\gamma$  because starting concentrations are small compared to concentrations of transcript in the presence of RS5444. qPCR Ct values for endogenous PDK4 are near the lower limit of the machine's sensitivity. This is not the case for DSCR1 which has relatively higher

**Figure 34: *DSCR1* is induced by RS5444 in MOSER S cells**

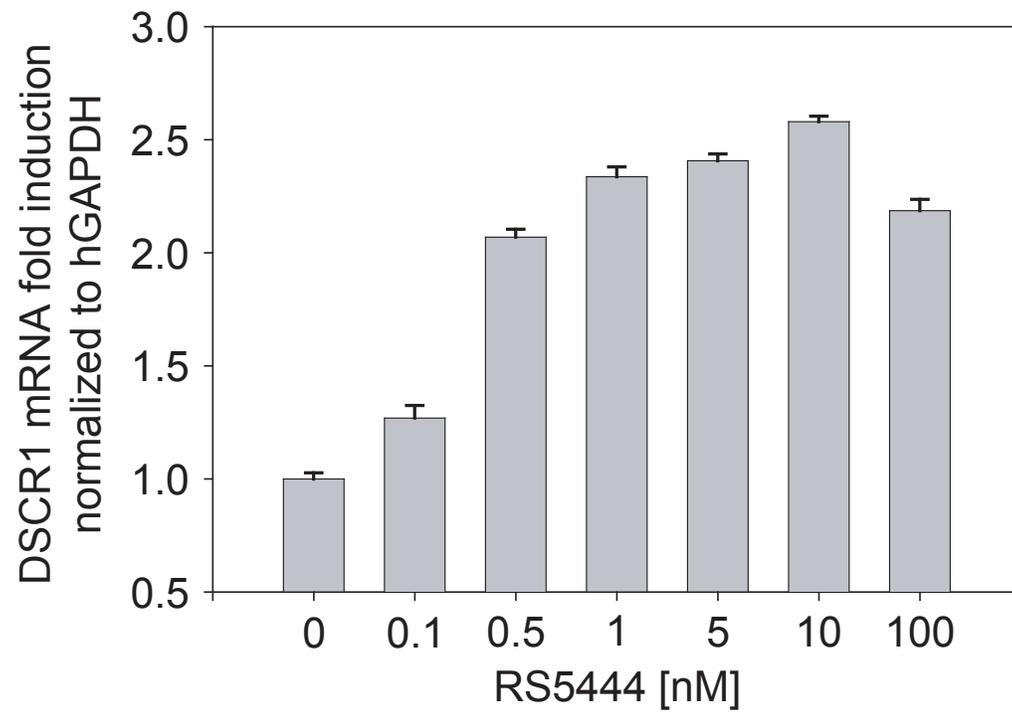
**A)** Total RNA was isolated every 6 h following RS5444 treatment. The abundance of *DSCR1* mRNA was measured by qPCR and normalized to that of human GAPDH. Data represents mean  $\pm$  SD, n = 3. **B)** HotellingT2 kinetic model of *DSCR1* from microarray time course every six hours in the presence and absence of RS5444. Green profiles represent triplicate RNA samples in the presence of RS5444. Red profiles represent triplicate RNA samples in the absence of RS5444.



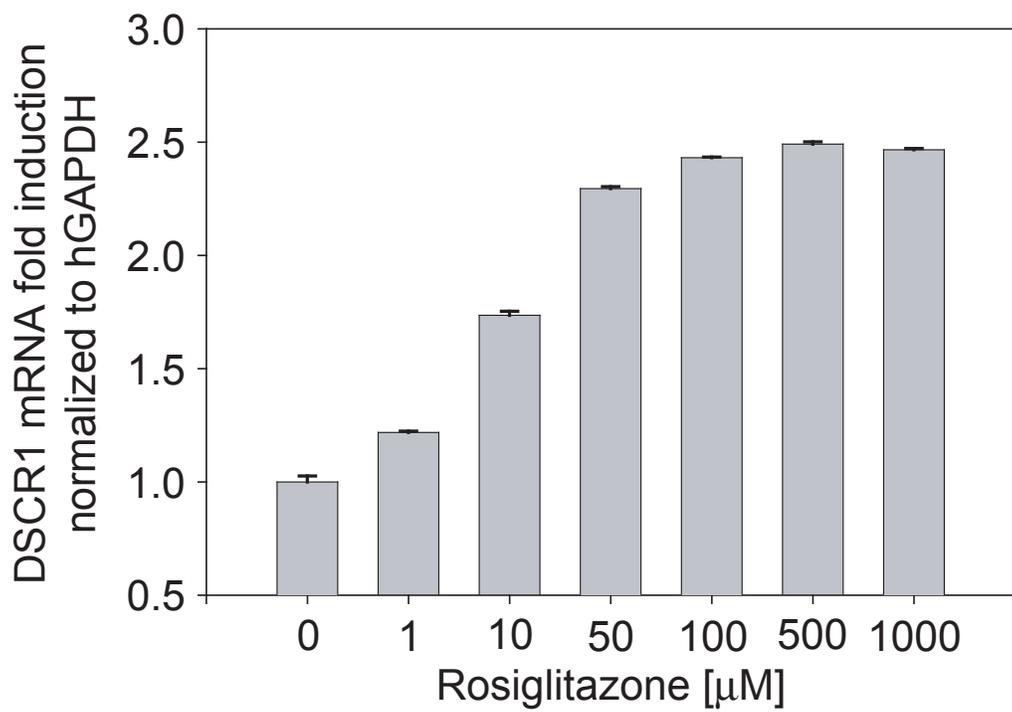
**Figure 35: *DSCR1* induction by thiazolidinediones is dose-dependent**

MOSER S cells were treated with varying concentrations of RS5444 (A) or rosiglitazone (B). The abundance of *DSCR1* transcript was measured by qPCR and normalized to that of human GAPDH. Data represent mean  $\pm$  SD, n = 3.

**A.**



**B.**



endogenous transcript abundance. Likewise, DSCR1 is not a primary target of PPAR $\gamma$  (see below). The additional requirement for DSCR1 transcription factor transcript production and translation may also account for this observation.

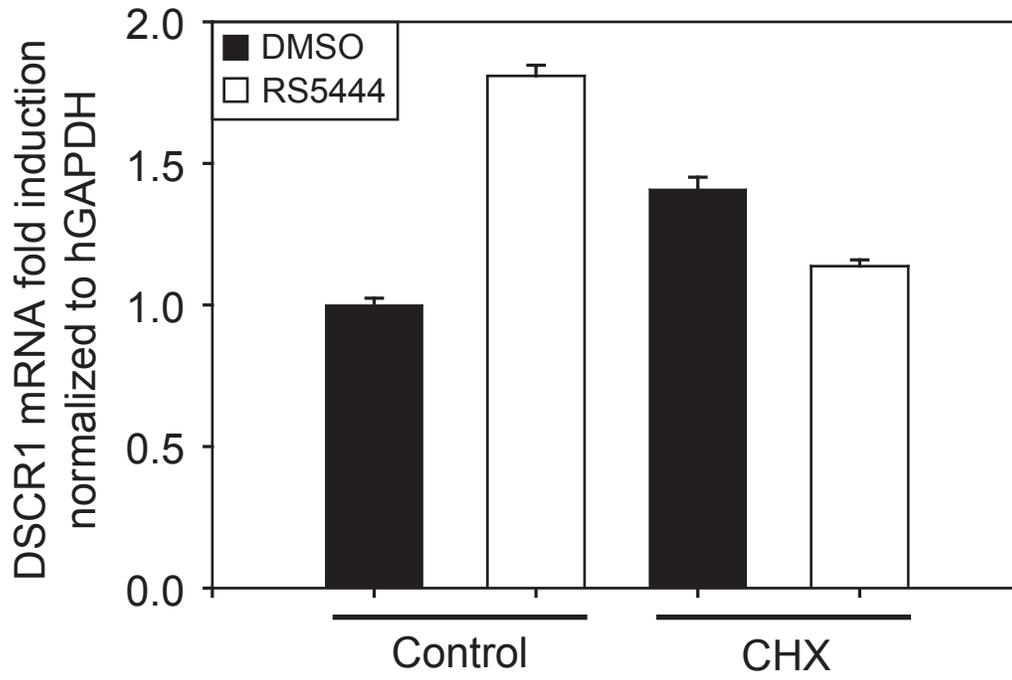
To test whether DSCR1 is a primary transcriptional target of PPAR $\gamma$ , we treated MOSER S cells with RS5444 in the presence or absence of cyclohexamide [Figure 36]. Cycloheximide is a general inhibitor of protein synthesis. Primary responsive genes will still be regulated in the presence of cycloheximide because their expressions are under the direct transcriptional control of endogenous PPAR $\gamma$ . However, once cycloheximide is added to the control, further protein production is halted. Therefore, any gene requiring the production of a transcriptional regulator (i.e. secondary target) for transcriptional control will not be regulated as a result of cycloheximide.

To test whether calcineurin and NFATc are required for DSCR1 induction by RS5444, MOSER S cells were treated with RS5444 in the presence or absence of cyclosporine A. Cyclosporin A is a well characterized calcineurin inhibitor. It binds to calcineurin thereby inhibiting mechanisms of calcineurin's activity, and ultimately abolishes nuclear localization of NFATc (204). Cyclosporine A blocks the induction of DSCR1 by RS5444, indicating that calcineurin and NFATc are required for DSCR1 induction [Figure 37A]. In contrast, cyclosporine A has no effect on the expression of PDK4, which is not regulated by the calcineurin-NFATc pathway [Figure 37B]. These data are consistent with what is previously known about NFATc regulation of DSCR1 (141, 205). These data also suggest that the induction of DSCR1 is PPAR $\gamma$  dependent, because cyclosporine A alone did not alter DSCR1 expression [Figure 37A].

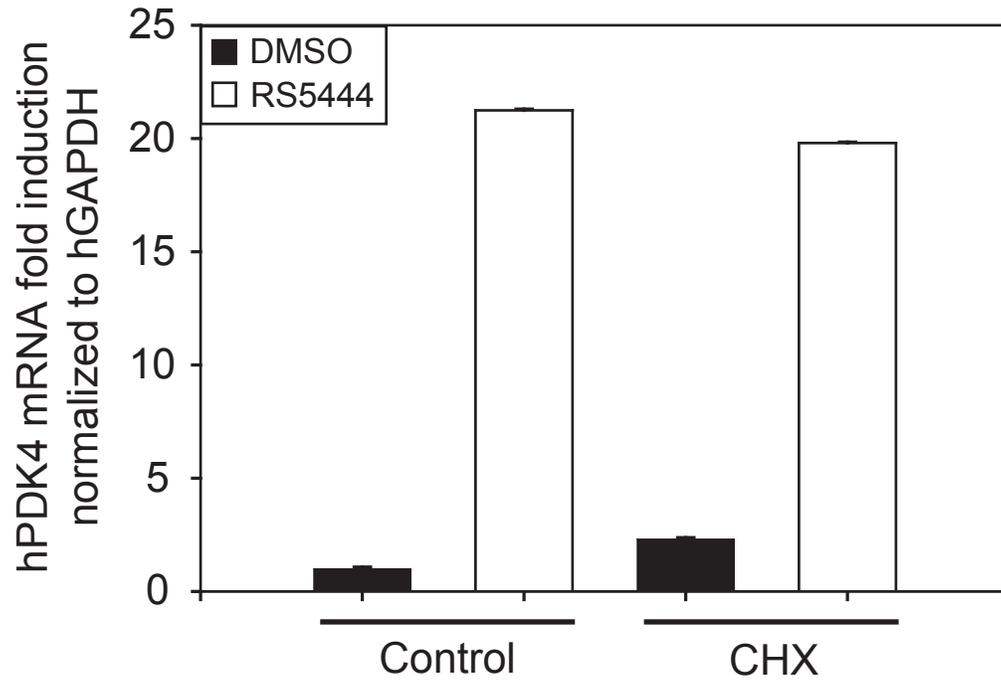
**Figure 36: *DSCR1* is a secondary target of *PPAR* $\gamma$**

MOSER S cells were treated with RS5444 in the presence and absence of 1 $\mu$ g/ml cycloheximide for 6 h. The abundance of *DSCR1* (**A**) or *PDK4* (**B**) mRNAs was measured by qPCR and normalized to that of human GAPDH. Data represent mean  $\pm$  SD, n = 3.

**A.**



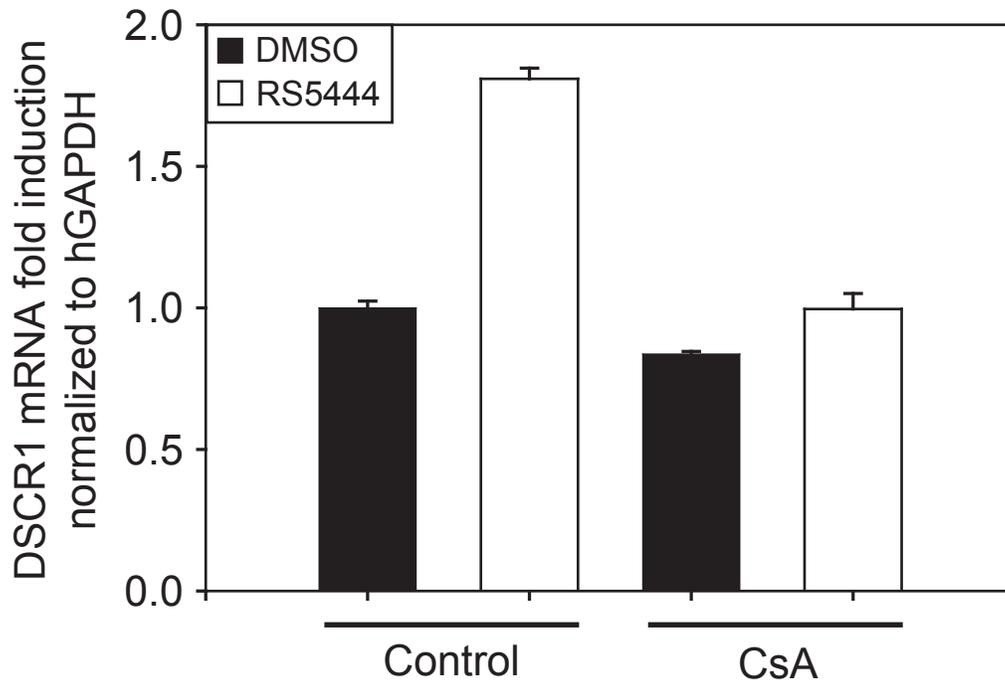
**B.**



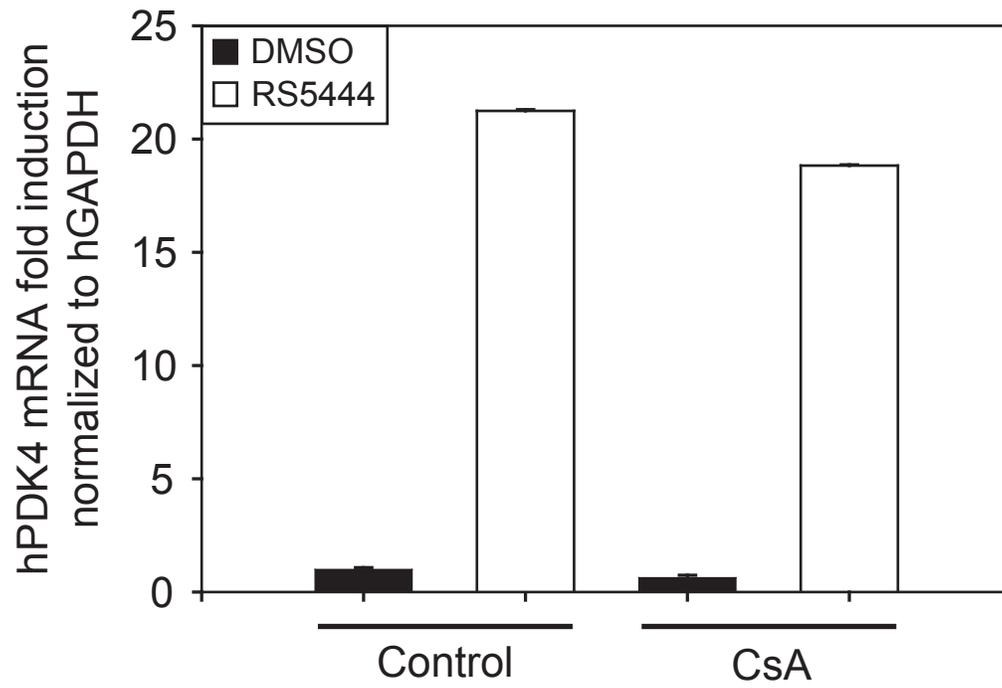
**Figure 37: Calcineurin activity is required for PPAR $\gamma$  induced DSCR1 expression**

MOSER S cells were treated with RS5444 in the presence and absence of 1 $\mu$ g/ml cyclosporine A for 6 h. The abundance of DSCR1 (**A**) or PDK4 (**B**) mRNAs was measured by qPCR and normalized to that of human GAPDH. Data represent mean  $\pm$  SD, n = 3.

A.



B.



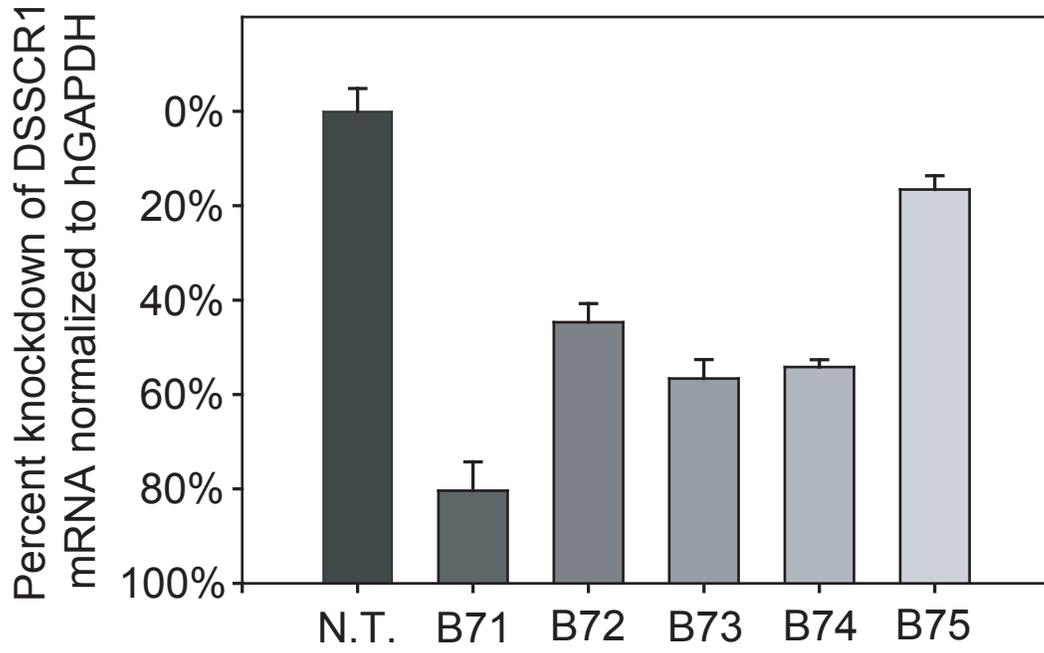
DSCR1 is a downstream target of calcineurin-NFATc pathway (205). However, DSCR1 is also a calcineurin inhibitor. Through a negative feedback loop, DSCR1 can block the calcineurin-NFATc pathway and subsequently block further induction of DSCR1 itself. We have shown that PPAR $\gamma$  induces DSCR1 and thus perturbs the basal calcineurin-NFATc-DSCR1 signaling pathway in MOSER S cells. To study whether the calcineurin-NFATc-DSCR1 pathway is involved in PPAR $\gamma$  mediated suppression of MOSER S cell proliferation and invasiveness, we used short hairpin RNA lentivirus techniques to silence the DSCR1 gene as discussed in Material and Methods. Five short hairpin RNA sequences were designed against different regions on the DSCR1 gene [Table 2]. MOSER S cells were transiently infected with lentivirus carrying these five candidate shRNAs. Of the five candidates, sequence B71 was the most effective in silencing the DSCR1 gene. Sequence B71 knocked down DSCR1 expression by more than 80% [Figure 38A] and its knockdown efficiency was relatively stable over a five day period [Figure 38B]. We established a culture of MOSER S cells infected with the B71 shRNA (referred to as B71 in the following discussion). In addition to inhibit basal DSCR1 expression, the shRNA also abolished RS5444 induced DSCR1 expression in B71 [Figure 39].

We then studied both the proliferation [Figure 40A] and invasion [Figure 40B] phenotypes of MOSER S/B71 using MOSER S infected with non-target control shRNA lentivirus as a control. Our data showed that RS5444 was not able to suppress the proliferation of MOSER S/B71, as shown in Figure 40A. Likewise, RS5444 was not able to inhibit MOSER S/B71 invasiveness either [Figure 40B]. These data suggest that

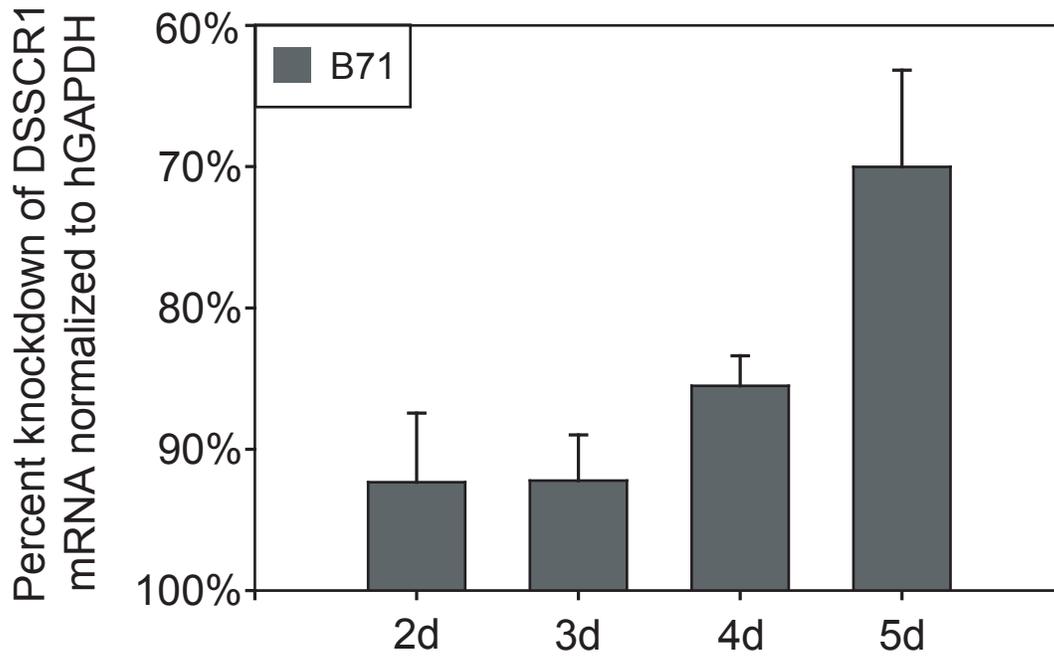
**Figure 38: *Characterization of DSCR1 sh-RNA lentiviral candidates***

MOSER S cells were infected with five different lentiviral sh-RNA candidate sequences that were designed to silence the DSCR1 gene. From **Table 2**: B71 = TRC#19844. B72 = TRC#19845. B73 = TRC#19846. B74 = TRC#19847. B75 = TRC#19848. Cells infected with a non-target lentivirus was used as control. Total RNA was isolated 48 h post infection. DSCR1 transcript abundance was measured by qPCR and normalized to that of human GAPDH. Data represents mean  $\pm$  SD, n = 3.

A.

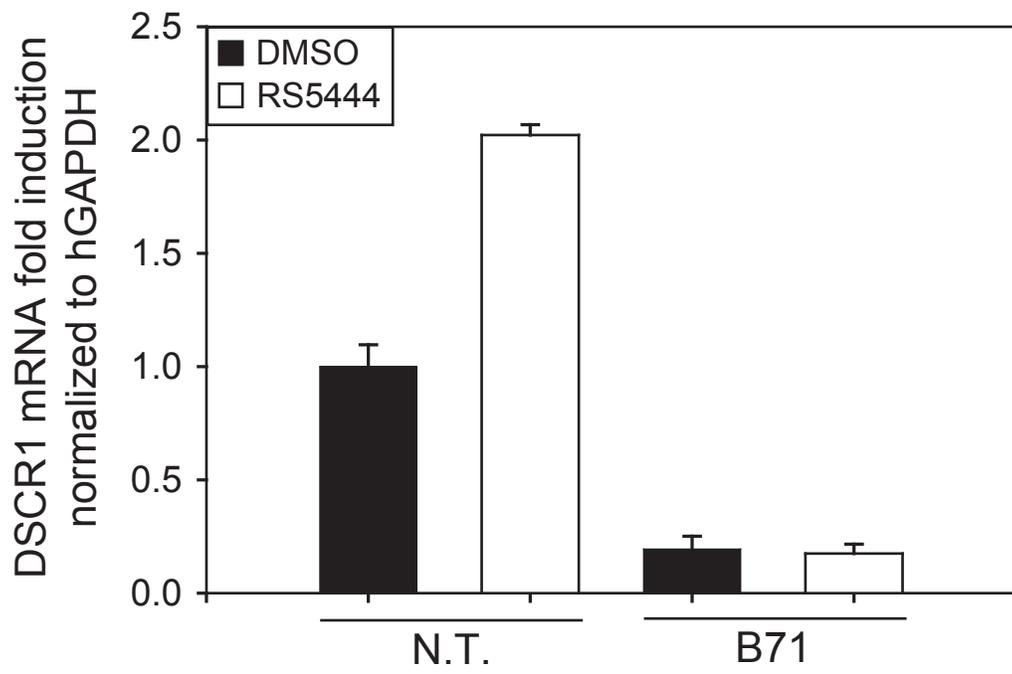


B.



**Figure 39: *DSCR1 sh-RNA B71 inhibits DSCR1 induction by PPAR $\gamma$***

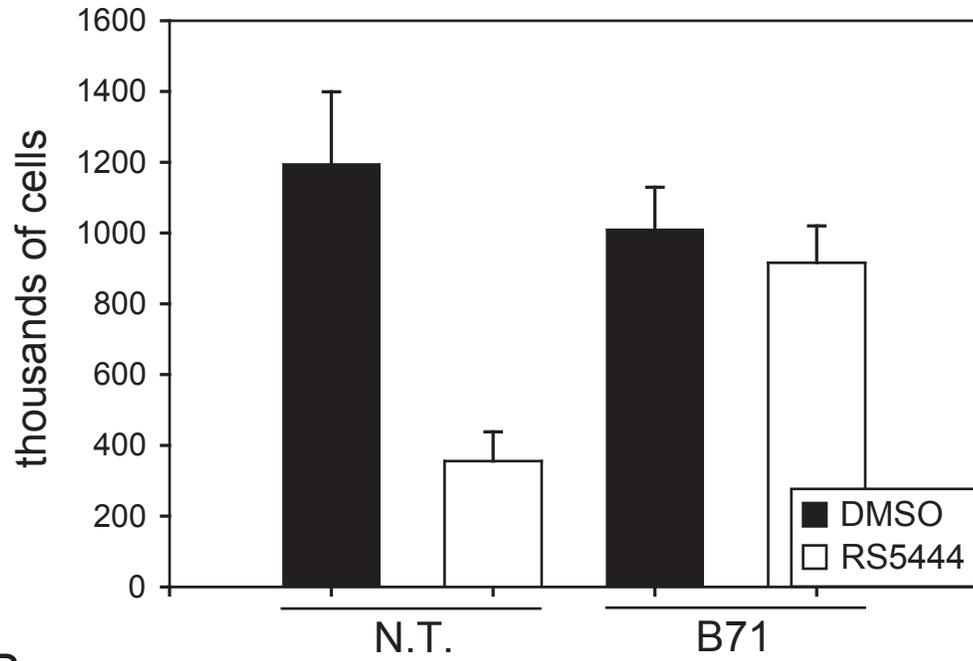
MOSER S/B71 and the non-target control MOSER S cells were treated with or without RS5444 for 24 h. *DSCR1* transcript abundance was measured by qPCR and normalized to human GAPDH. Data represents mean  $\pm$  SD, n = 3.



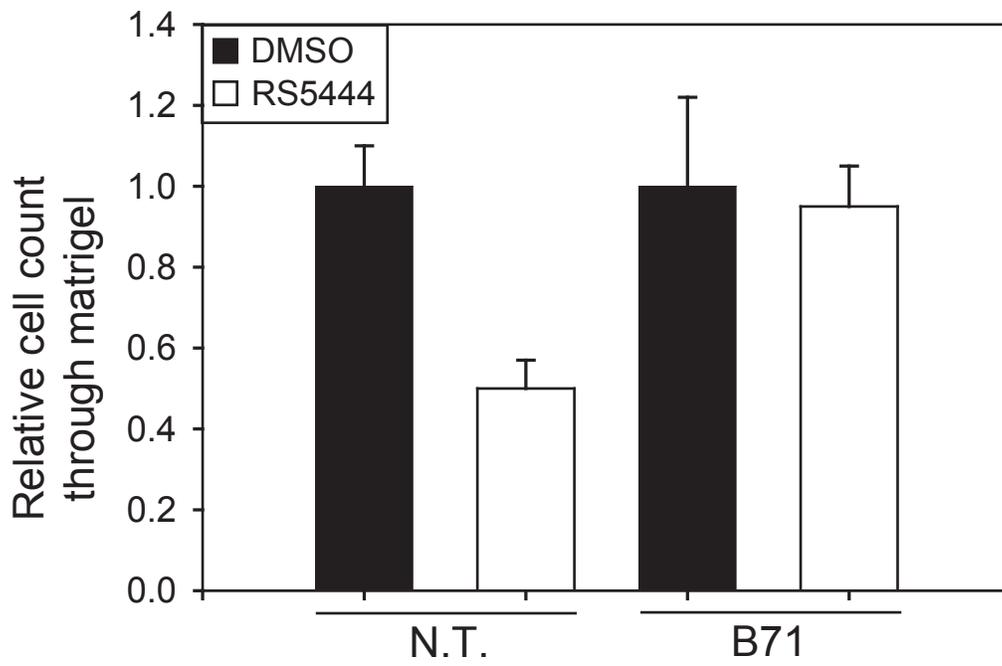
**Figure 40: *Silencing of DSCR1 gene blocks PPAR $\gamma$  induced suppression of MOSER S cell proliferation and invasion***

**A)** MOSER S/B71 and the non-target control MOSER S cells were treated with RS5444 for 5 days. Cells were counted with Beckman coulture counter. Data represents mean  $\pm$  SD, n = 3. **B)** MOSER S/B71 and the non-target control MOSER S cells pre-treated with RS5444 for 24 h were loaded onto matrigel covered invasion chamber and allow to invade for 18 h. Cells attached to the other side of chamber membrane were stained with crystal violet and counted under microscope. Data represents mean  $\pm$  SD, n = 6

A.



B.



PPAR $\gamma$  induced suppression of MOSER S cell proliferation and invasion is dependent on and mediated by the calcineurin-NFATc-DSCR1 signaling pathway.

## CHAPTER 4: DISCUSSION

Extensive research has been performed on the role of PPAR $\gamma$ 's regulatory effects in colorectal cancer, and while the corpus of knowledge from such studies has been substantial, the majority of experiments have been approached from a traditional, reductionist point of view. This is to say, the overall experimental design was hypothesis driven and contingent on prior knowledge of the system under investigation. This dissertation focuses upon elucidation of the genomic consequences of activated PPAR $\gamma$  from a holistic, systems biology perspective. The intent, following the system's biology paradigm, was to allow the data to dictate the next iteration of experiments to be performed. This was accomplished by first conducting a core set of experiments to characterize the phenotypic response of MOSER S colorectal cancer cells treated with a novel PPAR $\gamma$  ligand. MOSER S cells were chosen as our model system because others have demonstrated this cell line is sensitive to the TZD rosiglitazone (183). Likewise, it is widely known that activation of the nuclear hormone receptor PPAR $\gamma$  by TZD's in several cell models induces a large network of genetic response (126, 183, 206-208), and who's response has been linked to aspects of both growth arrest and invasion many cell types (209-213).

Based on these observations, we reasoned that given enough functional genomic time course data and the latest bioinformatic and systems biology techniques and technologies, it would be possible to map and analyze the genomic consequences of activated PPAR $\gamma$  in much finer detail than previously possible by the research

community. These microarray data were processed and analyzed using a unique protocol developed by our laboratory. Through the use of several sophisticated statistical and bioinformatic methods, a novel pathway was identified which offers an explanation for the phenotypic responses we observed in treated MOSER S cells. The pathway model, in turn, provided us with several hypotheses to test and confirm. At a fundamental level, we believe this unique discovery driven approach offered an unprecedented look into the unique role activated PPAR $\gamma$  plays in the genomic signaling of colon cancer. The results of our analyses have defined a novel signaling pathway that may have considerable significance in terms of our long term goals to develop PPAR $\gamma$  agonists as colon cancer chemotherapeutic or chemopreventive agents.

We began our study by characterizing the bioactivity of the novel TZD RS5444. We determined that RS5444 is a PPAR $\gamma$  specific agonist that activated the PPAR $\gamma$  primary target PDK4 at a dosage 1/20<sup>th</sup> that of the more commonly used TZD rosiglitazone, with an estimated EC50 roughly equivalent to 10 nM. This single observation is of primary importance to this study, for this concentration of agonist offers a certain degree of protection from unwanted activation of off-target effects which others have observed using higher doses of TZD's (214).

MOSER S cells proved to be an excellent choice as our model cell system. Because of their sensitivity to TGF $\beta$  and are mutated of APC function within the prototypical loss-of-function mutation (premature stop codon at exon 15), we concluded that this cell line offered the best circumstances to study the effects of PPAR $\gamma$  ligands as a potentially exceptional model of mildly transformed colorectal cancer cells. Activated

PPAR $\gamma$  in MOSER S cells did not regulate its own receptor abundance at either the transcript or protein level and offers an additional intrinsic control for time course studies. Furthermore, we observed profound growth arrest of MOSER S cells culture in the presence of 10 nM within the first three days. This observation is consistent with the observations of others culturing human colorectal cancer cells in the presence of rosiglitazone, including HT-29 cells (208) and MOSER S cells (183).

We likewise observed growth arrest of MOSER S cells cultured in the presence of RS5444 for several weeks in soft-agar. The results were obtained using Kodak camera imaging and software to count the relative coverage of colonies in both treated and untreated cultures (see Materials and Methods). This technique allowed for a much larger initial seeding of colonies than is traditional employed. We reasoned that the larger number of initially seeded colonies increased the sensitivity of the test. Granted, this technique assumes the area of coverage represents monolayers of cells, which is simply not the case in a three dimensional soft-agar colony. However, we believe the large number of colonies mitigate this effect, and in the end, the large number of cells offsets any that may have been photographed overlaying one another. A large majority of the colonies survived the three weeks cultured in the presence of RS5444 (data not shown) and therefore, the anti-proliferative effects under anchorage-independent conditions appear to be largely the result of blocking mitosis and not signaling apoptosis. We did not conduct tests specific for measuring either of these cases, and further work will be required to elucidate the exact cause of these results.

The effects on MOSER S cells cultured in the presence of RS5444 appear to be irreversible. These data suggest that activation of PPAR $\gamma$  in MOSER S cells by RS5444 caused a subset of cells to permanently stop proliferation. Irreversible withdrawal from the cell cycle is one of the hallmarks of differentiation, and our data suggest that PPAR $\gamma$  may cause differentiation of some MOSER S cells. The exact mechanism of action remains unclear. Others have demonstrated that activation of PPAR $\gamma$  may cause either G1-arrest (206, 208) or apoptosis (215) in HT-29 cells. We did not investigate this issue further and it is clear that additional work will be required to explain the precise mechanisms of action responsible for these observations.

Finally, and perhaps the most interesting result of these characterization studies, was the observation that activation of PPAR $\gamma$  appears to prohibit invasion, but not motility in MOSER S cells, by blocking the cell's ability to digest matrigel (216), but its locomotive ability through porous material. This observation is rare and several interesting candidate scenarios may provide explanation for these findings, most of which involve gene expression control important in angiogenesis and GTPase activity (68).

These results clearly showed that at the very least several rather interesting pathways were under the control of activated PPAR $\gamma$ . Our initial characterization experiments provided evidence that a functional genomics time course experiment run over a twenty-four hour period after the addition of agonist would be sufficient to capture most if not all of the genomic consequences from PPAR $\gamma$  activation. However, it became clear early in the project that proper experimental design and clearly defined

assumptions would be paramount to gain the most information from microarray data. Compounded with the initial effort and resulting expense associated with these experiments, understanding the biological system is necessary to make reasonable estimates of results and improve one's chance of a successful outcome.

Tai and Speed categorize some of the earlier efforts predicting differentially expressed genes as a function of both biological stimulus and time using microarrays (217). Specifically, Tai and Speed discuss the particular strengths and weaknesses of both one-sample, two-sample and multi-sample analysis of such efforts in the context of longitudinal and cross-sectional experimental design. Briefly, longitudinal microarray datasets are comprised of mRNA isolated from the same experimental *unit* or biological sample where each sample represents one ordinate value per variable (gene) and may be assembled by any number of methods into a discrete or continuous temporal kinetic profile with an associated correlation and covariant relationship. Cross-sectional microarray datasets, on the other hand, are formed from isolated mRNA samples across time from different experimental units. That is, cross-sectional datasets allow the additional sample relationship across a particular time point, such as using average transcript expression values to represent the population pool across several replicates. Ultimately, depending upon experimental design and limitations of biological samples, the resultant dataset may have characteristics of both experimental designs mentioned above and care must be taken in both the choice of analytical models and one's starting assumptions when interpreting results.

Given this information, it was decided to utilize a two-class longitudinal array experimental design. That is, technical biological replicates (replicate samples from the same parental cell line clone) were isolated at given time points in the presence and absence of a biological stimulus (nuclear hormone analogue). A stock of MOSER S cells was cultured and expanded to accommodate an initial seeding of all culture plates at once. This is, of course, different than repeating the time course experiment three different times, using one replicate for each group and time point. We decided to use technical replicates to decrease the amount of biological noise that ultimately would have been introduced into the microarray data analysis. We likewise felt the amount of information loss that might occur from designing the experiment in this fashion was significantly outweighed by the reduction of noise and resultant operator error from conducting three challenging 24 h time course experiments.

Once the Affymetrix microarrays were constructed and passed our various quality controls, we focused our efforts on determining the best preprocessing method for obtaining expression data. Three different methods were tested: MAS5, RMA, and GCRMA using the Bland-Altman (a.k.a. MvA plots) test. Bland-Altman plots provided a straightforward method to compare the performance of one test versus another test. In this case, we wanted to examine the reproducibility between one microarray replicate at a particular time point versus another microarray replicate at the same time point.

The results are depicted as a scatter plot, where each datum represents a pair-wise comparison of each probe set between the two microarray chips in question. The x-axis represents the averages of expressions for each probe set between the two replicates over

the entire range of expression. The y-axis represents the differences in expression over the entire range of expressions. A lowess regression line can then be drawn through the data points. A score can also be assigned to the lowess line, where higher numbers denote more dissimilarity. In this context, more dissimilar scores reflect a model's inability to provide similar replicates. So, for example, if there is no difference between two microarray replicates, the resulting lowess line would be flat, passing through the origin, with a resulting dissimilarity score of zero.

Our analysis using Bland-Altman plots produced rather interesting results. The most obvious was that GCRMA proved to be a superior preprocessing and summarization method, based on its lowess regression score. What we did not expect from this analysis was the rather poor performance of MAS5 relative to the other two tests, as well as raw-unadulterated data. However, these data are generally consistent with performance tests of summarization conducted by others (72). We believe our results are indicative of the default settings we used for bioconductor's MAS5 implementation. The default settings for MAS5 provide no normalization. Likewise, we did not expect to observe such odd skewing from RMA preprocessing on the 24 h control. This was not observed for all cases of RMA (data not shown).

We then turned to statistical bootstrapping for an objective measurement of performance given several models for determining differential expression. Our results show that the model-based empirical Bayes algorithm outperformed both the non-parametric algorithm as well as the polynomial regression algorithm. These results, however, must be taken in context. This bootstrapping method above does prove that

HotellingT2 outperforms the other two algorithms in statistical power, but only for this particular dataset. Essentially, the entire longitudinal microarray dataset represents one observation, and therefore, the statistical conclusions drawn can only be understood in the context of this one observation. For a more thorough statistical analysis of differential expression performance, the entire longitudinal microarray experiment would need to be performed several times, ideally by different researchers following our protocol. Only with this huge, and ultimately cost-prohibitive, collection of data could more definitive statistical claims be made. However, given our restraints, we feel confident that the empirical Bayes HotellingT2 output provides a very good first approximation of the functional genomics in play under the control of activated PPAR $\gamma$ .

Several broad conclusions can be drawn from the descriptive statistics of these data. First, most genes are down-regulated. Only 133 (about 7%) of the 1975 Affymetrix probe sets were up-regulated. Most of these probe sets map to ontology functions involved in metabolism (data not shown). This is to be expected, given the mechanisms of action TZD's play in lipid metabolism (218). Secondly, the kinetic profiles of high scoring probe sets could be roughly classified into four categories. Genes that were modeled as up-regulated either reach saturation or did not reach saturation and continued increasing in abundance over the twenty-four hour testing period. The same was observed for down-regulated genes. Additionally, within these four classes, some probe sets were modeled in which the control groups (i.e. in the absence of RS5444) changed over time, but treated groups (i.e. cultured in the presence of RS5444) changed. Our current understanding of RS5444's mechanisms of action provides no explanation for

these observations. It may be the case that we observed probe sets whose representative genes are somehow receptive to DMSO, which would otherwise remain unchanged in the presence of RS5444. Unfortunately, our tests were not designed to further elucidate this possibility and these additional tests will need to be administered.

One other aspect of the empirical Bayes HotellingT2 model is worth mentioning. The model is biased towards large changes in individual transcript abundance. The model scores probe sets that change more in treated versus control (and the other way around) higher than those that do not. The score increases with greater change temporally or spatially or both. Therefore, this model biases the effects of change as most important with no consideration to the underlying biological context of the represented probe sets. The does not take into account synergy of, for example, a small group of genes with low overall transcriptional change in abundance, but who's combined effect is sufficient to provide a genetic switch of some sort. Those kinds of details remain hidden and encoded within hierarchy of HotellingT2, if they make the cut-off at all. Indeed, much work remains in bioinformatic software engineering to include evidence-based prior probabilities of gene interaction into models of transcriptional relevance.

This point was relevant in our use of gene-ontology software. GO analysis was initially used to check our HotellingT2 results against what had previously been published. As expected, GO super classes involving proliferation, metabolism, and migration/motility were all scored very high relative to others. However, hidden within the list of classes that achieved a statistically-significant cutoff was calcium-mediated

signaling; of which the provided the basis for the conclusion of this project. Moreover, while this class did not score as high as others, it contained one of the highest scoring (and ultimately most important) genes within the realm of this project: DSCR1. Granted, calcium-mediated signaling did not score as well as, say, proliferation because the way GO overrepresentation analysis is coded. Yet, it remains unclear from these data how one would code software to pull out the DSCR1 target, when so little has been published on DSCR1's role within the context of colorectal cancer. Moreover, as far as we are aware, nothing has been published on DSCR1 in the context of PPAR $\gamma$ -mediated regulation.

And therein lies the challenge involving the current state of affairs of systems biology. For now, systems biology's greatest strength and promise (holistic, grand-context research) is its own largest drawback. There appears to be a critical mass, a threshold of data, required to contextualize programs and algorithms for *a priori* molecular target identification. Clearly the community is not there yet. In the meantime, through much literature searching, we ultimately came to the conclusion that DSCR1 and its role in calcium-mediated signaling offered a potentially excellent explanation for the phenotypic responses we observed *in vivo*.

Ingenuity Pathway's Analysis provided us with the necessary bridge to strengthen our newly formed hypothesis that DSCR1, and in a larger context calcium-mediated signaling, was responsible for functions involved in proliferation and invasion under the transcriptional control of PPAR $\gamma$ . IPA inferred the transcriptome involving the negative-feedback loop between the transcription factor NFATc and DSCR1 and downstream

genes involved in the tumor suppressor properties of PPAR $\gamma$ . This conclusion appears to hold for other colorectal cancer cell lines sensitive to PPAR $\gamma$ . For example, we found a correlation between the abundance of DSCR1 transcript and the ability of PPAR $\gamma$  to inhibit proliferation in colorectal cancer cell lines (unpublished data from our laboratory). Likewise, of the four cell lines we tested, MOSER and HT-29 cells are the most sensitive to PPAR $\gamma$  (unpublished Affymetrix genechip data from our laboratory), and likewise have the highest fold induction of DSCR1 by PPAR $\gamma$  (unpublished data from our laboratory).

Upon knockdown DSCR1's transcript abundance by sh-RNA lentiviral particles, we observed ablation of both proliferation inhibition and invasion in MOSER S cells. In essence, DSCR1 appears to be a master regulator of NFATc controlled transcriptional regulation by mediating the bioactivity of calcineurin—a serine/threonine protein phosphatase required by NFATc for nuclear translocation. Therefore, as a result of knocking down DSCR1, one can override PPAR $\gamma$ 's ability to suppress proliferation and invasiveness.

Other questions logically stem from these observations and findings. For example, what is upstream of calcineurin that is regulated by PPAR $\gamma$ ? Are these genes important for controlling proliferation and invasion in colorectal cancer? One obvious target that needs more attention is the role the gene vascular endothelial growth-factor (VEGF) plays in this signaling axis. Our ontology analysis revealed angiogenesis as a statistically significant ontological function under PPAR $\gamma$  control. VEGF is an

interesting target to consider because others have shown a direct connection between VEGF signaling and DSCR1 (137). It is conceivable that up-regulation of VEGF, as modeled by our empirical Bayes HotellingT2 analysis, regulates calcium-dependent calcineurin activity through phosphatidylinositol 3 kinase (PI3K). Of course, follow-on experiments would need to verify the abundance and activity of the VEGF receptor along with the activity of a kinase cascade. Another interesting candidate one might consider is the top scoring gene in our empirical Bayes HotellingT2 model: Arrestin Beta 1 (ARRB1). ARRB1 is an inhibitor of G protein-coupled receptors through several pathways including angiotensin II type 1A (219, 220) and crosstalk with receptor tyrosine kinases (221). G-protein couple receptors promote calcium up-take. Therefore, it is conceivable that up-stream of calcineurin, the repression of ARRB1 by PPAR $\gamma$  is influencing NFATc through calcineurin.

As is often the case in a new evolving field of science, there are potentially many more methods which could be used with this unique data set offering further insight into the role PPAR $\gamma$  plays in colon cancer. A current area of intense research in systems biology involves data-integration from multiple data sources. For example, it is conceivable that one may repeat the time course experiment used in this dissertation and isolate conditioned media at the same time points, under the same conditions. Assuming enough conditioned media is acquired, it is possible to create one dimensional H<sup>+</sup> Nuclear Magnetic Resonance (NMR) spectra. 1-D NMR spectral data can be manipulated to represent abundance information on proteins and small molecules within the sample. These spectral data would then, assuming properly conducted experimental design, be

capable of capturing through time, molecular and protein markers important in paracrine signaling. Likewise, metabolic and waste markers may also be identified. This so-called metabolomic dataset could then be used in conjunction with the genomic data to provide an even higher, and more complex, view of PPAR $\gamma$ 's role in colorectal cancer cells. This technique has been used successfully in toxicology (222).

Likewise, an important hypothesis of clinical importance stems from these findings. First and foremost, DSCR1 expression within tumor tissue may need to be considered when clinicians attempt chemotherapeutic treatment involving PPAR $\gamma$  agonists. A lack of DSCR1 expression may explain why some have reported that, under certain circumstances and with certain tissues, PPAR $\gamma$  can act as a tumor enhancer. Our group has considered these possibilities in earnest. We have begun the requisite steps needed to test this hypothesis in DSCR1 knockout mice.

This dissertation is a deliberate recess from more traditional biological experimental methods and is in some ways a rather audacious attempt to take advantage of the untapped potential of the systems biology approach. Perhaps the most important difference between hypothesis driven research and discovery driven research is the set of starting assumptions. It is typical practice in hypothesis driven research to begin with a relatively simple model of the system. This model allows one to test a discrete, carefully defined hypothetical prediction derived from a well-articulated set of assumptions about how the system will respond to a specific perturbation. This is fundamentally different from systems biology projects which typically make no prior assumptions of the problem

at hand and relies heavily on manipulation of huge quantities of data to dictate the next step within an ever-changing experimental design.

Another important difference between the two methodologies is the use of different experimental technologies, each having their own strengths and weaknesses. Traditional biological experimentation benefits from well understood principles and relatively simple data analysis. Well-designed and carefully executed experiments of this sort have the ability to yield clear cut answers to very specific questions. Moreover, these methods benefit from generalization. Generalization in this context means that a single method is attractive to a wide research audience by providing utility to a wide set of problems. For example, western blots are possibly one of the widest used technologies in biology and are therefore by definition highly generalized. The ability to measure the abundance and size of proteins is attractive to a broad contingent of life science research. But perhaps most important, these techniques are straightforward and cost-effective. Indeed, a good portion of what we currently know about biochemical signaling was deduced using these approaches. However, at the most fundamental levels, these methods rely on the assumption that cause and effect relationships inherent in the experimental design are linear and binary, and it is becoming increasingly clear that biology often, perhaps invariably, derives from complex, non-linear interactions.

Systems biology techniques on the other hand are largely agnostic to the level of complexity implicit within the system. Unfortunately, while the theory of systems biology methods is compelling, the discipline suffers from a want of mature methodology. In practice, a new system's biology approach often provides utility to a

limited audience. For instance, it is often the case that a tool or algorithm is created that is quite good at solving one kind of problem. This is partly the manifestation of a limited number of publicly available datasets, which in turn, limits the researchers in the kinds of questions that can be answered. An obvious consequence from such limitations is a set of tools with reduced utility, consequent on those working with a limited set of data. In essence this problem produces a catch-22 situation. Lack of systems biology methodologies shy those away from producing the requisite data sets needed for advancement of the field. Further research is also confounded by technical challenges and the required foresight needed to produce these data. In summary, there clearly exists within the systems biology community a need for: 1) better, more generalized data sets with clearly defined experimental goals, and 2) protocols and methodologies that address these datasets.

The application of a systems biology approach therefore presents a challenge. It is widely believed that systems biology holds the key to the holy grail of modern experimental biology: the use of mathematical models and computer algorithms to predict pathways that control important biological processes. The challenge therefore is to develop the models and algorithms that allow one to reliably predict the structure of the data and to integrate such information into an informatic framework that facilitates the development of discrete testable hypotheses. This challenge is tempered by the necessity of appropriate data sets that address clearly defined goals and yet are broad enough in scope to serve a clear purpose to the systems biology community.

This dissertation is an attempt to take us one step closer to realizing this goal. This was accomplished by integrating the major tenets of systems biology into a simple recursive recipe: 1) identify and characterize a system for the task; 2) perturb the system and measure system-wide responses; 3) manage and analyze these data; 4) use these data for the next iteration of experimentation. The weakest link in this recursive work flow, as mentioned before, is the tools required to analyze and make sense of the data to provide a road map to the next step. In order to resolve this problem, one must first define the biological components of the system in step 1: identification and characterization of a system for the task. These components would include at the highest levels of abstraction gene structure and function, protein expression and activity, and the products of these proteins and the activity of these products. The disciplines that have evolved around these three components are respectively genomics, proteomics, and metabonomics. These three disciplines focus largely on aspects of the second step within the recursive recipe: perturbing the system and measuring the response. The third step is to establish some way to analyze the interactions between these components including techniques borrowed from computer science, biostatistics, and bioinformatics. This dissertation follows this recursive recipe and provides a clearly defined dataset and a clearly defined question to be answered. We believe the data set and results provide the necessary next step in the systems biology community's realization of its goal.

## APPENDIX A

Probe Set	Gene Title	Gene Symbol	Hotellin gT2 Scores	MB Statistic	Max FC
203757_s_at	carcinoembryonic antigen-related cell adhesion molecule 6 (non-specific cross reacting antigen)	CEACAM6	3958.1	-0.19024	4.48
211883_x_at	carcinoembryonic antigen-related cell adhesion molecule 1 (biliary glycoprotein)	CEACAM1	2671.2	-0.32025	3.81
209498_at	carcinoembryonic antigen-related cell adhesion molecule 1 (biliary glycoprotein)	CEACAM1	4892.8	-0.13520	3.68
212099_at	ras homolog gene family, member B	RHOB	3225.2	-0.25306	3.67
224566_at	trophoblast-derived noncoding RNA	TncRNA	1230.8	-0.70872	3.63
241940_at	Abhydrolase domain containing 3	ABHD3	1478.0	-0.59923	3.22
213816_s_at	met proto-oncogene (hepatocyte growth factor receptor)	MET	1773.1	-0.50161	3.06
211657_at	carcinoembryonic antigen-related cell adhesion molecule 6 (non-specific cross reacting antigen) /// carcinoembryonic antigen-related cell adhesion molecule 6 (non-specific cross reacting antigen)	CEACAM6	2008.3	-0.44104	2.93
201650_at	keratin 19	KRT19	1591.0	-0.55841	2.91
209122_at	adipose differentiation-related protein	ADFP	2830.4	-0.29860	2.88
212531_at	lipocalin 2 (oncogene 24p3)	LCN2	1160.2	-0.74651	2.80
214581_x_at	tumor necrosis factor receptor superfamily, member 21	TNFRSF21	2159.3	-0.40805	2.77
202672_s_at	activating transcription factor 3	ATF3	1516.1	-0.58495	2.74
225239_at	CDNA FLJ26120 fis, clone SYN00419	---	1768.7	-0.50288	2.71
208370_s_at	Down syndrome critical region gene 1	DSCR1	2256.1	-0.38886	2.69
209365_s_at	extracellular matrix protein 1	ECM1	1792.4	-0.49617	2.62
212806_at	KIAA0367	KIAA0367	1586.6	-0.55990	2.58
223541_at	hyaluronan synthase 3	HAS3	1897.8	-0.46795	2.57
217173_s_at	low density lipoprotein receptor (familial hypercholesterolemia)	LDLR	1140.0	-0.75799	2.51
202842_s_at	DnaJ (Hsp40) homolog, subfamily B, member 9	DNAJB9	997.2	-0.84898	2.44
1566472_s_at	all-trans-13,14-dihydroretinol saturase	RetSat	1708.5	-0.52063	2.39
210517_s_at	A kinase (PRKA) anchor protein (gravin) 12	AKAP12	1278.5	-0.68501	2.32
209457_at	dual specificity phosphatase 5	DUSP5	1249.3	-0.69935	2.22
215111_s_at	TSC22 domain family, member 1	TSC22D1	1528.2	-0.58049	2.20
201012_at	annexin A1	ANXA1	936.6	-0.89379	2.13

207401_at	prospero-related homeobox 1	PROX1	317.7	-1.86990	2.11
201559_s_at	chloride intracellular channel 4	CLIC4	883.8	-0.93649	2.09
244804_at	Sequestosome 1	SQSTM1	803.8	-1.00885	2.08
200710_at	acyl-Coenzyme A dehydrogenase, very long chain	ACADVL	1007.9	-0.84148	2.04
234989_at	trophoblast-derived noncoding RNA	TncRNA	1047.0	-0.81516	2.04
226817_at	desmocollin 2	DSC2	595.5	-1.25775	2.02
216594_x_at	aldo-keto reductase family 1, member C1 (dihydrodiol dehydrogenase 1; 20-alpha (3-alpha)-hydroxysteroid dehydrogenase)	AKR1C1	790.0	-1.02240	2.01
218856_at	tumor necrosis factor receptor superfamily, member 21	TNFRSF21	991.0	-0.85338	2.00
221060_s_at	toll-like receptor 4 /// toll-like receptor 4	TLR4	929.8	-0.89910	1.99
211026_s_at	monoglyceride lipase /// monoglyceride lipase	MGLL	722.1	-1.09428	1.98
224565_at	trophoblast-derived noncoding RNA	TncRNA	798.7	-1.01382	1.98
231973_s_at	anaphase promoting complex subunit 1	ANAPC1	553.4	-1.32310	1.98
209699_x_at	aldo-keto reductase family 1, member C2 (dihydrodiol dehydrogenase 2; bile acid binding protein; 3-alpha hydroxysteroid dehydrogenase, type III)	AKR1C2	720.3	-1.09626	1.94
209435_s_at	rho/rac guanine nucleotide exchange factor (GEF) 2	ARHGEF2	1279.8	-0.68440	1.93
201471_s_at	sequestosome 1	SQSTM1	905.4	-0.91854	1.93
213953_at	keratin 20	KRT20	962.1	-0.87442	1.90
226084_at	microtubule-associated protein 1B	MAP1B	669.5	-1.15689	1.90
218651_s_at	La ribonucleoprotein domain family, member 6	LARP6	494.3	-1.42717	1.88
1561775_at	---	---	402.2	-1.62684	1.87
202722_s_at	glutamine-fructose-6-phosphate transaminase 1	GFPT1	799.1	-1.01336	1.86
204151_x_at	aldo-keto reductase family 1, member C1 (dihydrodiol dehydrogenase 1; 20-alpha (3-alpha)-hydroxysteroid dehydrogenase)	AKR1C1	539.6	-1.34602	1.85
201464_x_at	v-jun sarcoma virus 17 oncogene homolog (avian)	JUN	944.2	-0.88791	1.85
209136_s_at	ubiquitin specific peptidase 10	USP10	507.9	-1.40169	1.84
221841_s_at	Kruppel-like factor 4 (gut)	KLF4	393.2	-1.64967	1.83
212233_at	Microtubule-associated protein 1B /// Homo sapiens, clone IMAGE:5535936, mRNA	MAP1B	465.7	-1.48353	1.83
224917_at	microRNA 21	MIRN21	583.7	-1.27535	1.83
227475_at	forkhead box Q1	FOXQ1	588.9	-1.26762	1.83
1553105_s_at	desmoglein 2	DSG2	778.1	-1.03431	1.81
224725_at	mindbomb homolog 1 (Drosophila)	MIB1	396.1	-1.64220	1.78

211708_s_at	stearoyl-CoA desaturase (delta-9-desaturase) /// stearoyl-CoA desaturase (delta-9-desaturase)	SCD	816.0	-0.99716	1.73
213017_at	abhydrolase domain containing 3	ABHD3	320.8	-1.85962	1.72
201702_s_at	protein phosphatase 1, regulatory subunit 10	PPP1R10	311.7	-1.89009	1.71
207535_s_at	nuclear factor of kappa light polypeptide gene enhancer in B-cells 2 (p49/p100)	NFKB2	649.9	-1.18213	1.68
228568_at	GRINL1A combined protein	Gcom1	681.1	-1.14256	1.67
215034_s_at	transmembrane 4 L six family member 1	TM4SF1	723.6	-1.09254	1.67
226886_at	Clone 114 tumor rejection antigen	---	351.1	-1.76511	1.65
202923_s_at	glutamate-cysteine ligase, catalytic subunit	GCLC	517.1	-1.38507	1.65
213986_s_at	chromosome 19 open reading frame 6	C19orf6	305.9	-1.91024	1.64
202793_at	putative protein similar to nessy (Drosophila)	C3F	759.2	-1.05384	1.61
241418_at	Hypothetical LOC344887	---	571.7	-1.29383	1.61
202067_s_at	low density lipoprotein receptor (familial hypercholesterolemia)	LDLR	374.2	-1.69976	1.60
208290_s_at	eukaryotic translation initiation factor 5	EIF5	732.2	-1.08301	1.59
221577_x_at	growth differentiation factor 15	GDF15	732.7	-1.08237	1.59
200615_s_at	adaptor-related protein complex 2, beta 1 subunit	AP2B1	659.4	-1.16971	1.56
210425_x_at	golgi autoantigen, golgin subfamily a, 8B	GOLGA8B	294.6	-1.95066	1.55
225698_at	TIGA1	TIGA1	789.9	-1.02248	1.54
209387_s_at	transmembrane 4 L six family member 1	TM4SF1	377.3	-1.69142	1.54
228284_at	transducin-like enhancer of split 1 (E(sp1) homolog, Drosophila) /// hypothetical gene supported by BC000228; NM_005077	TLE1 /// LOC389863	425.1	-1.57197	1.53
202887_s_at	DNA-damage-inducible transcript 4	DDIT4	777.3	-1.03511	1.52
214594_x_at	ATPase, Class I, type 8B, member 1	ATP8B1	380.0	-1.68409	1.51
228959_at	CDNA clone IMAGE:5262734	---	359.4	-1.74094	1.51
222802_at	endothelin 1	EDN1	358.7	-1.74307	1.50
202669_s_at	ephrin-B2	EFNB2	485.0	-1.44501	1.49
207275_s_at	acyl-CoA synthetase long-chain family member 1	ACSL1	492.8	-1.42998	1.49
230256_at	Chromosome 1 open reading frame 104	FLJ35976	325.1	-1.84560	1.48
238666_at	Farnesyl-diphosphate farnesyltransferase 1	FDFT1	674.8	-1.15037	1.48
224354_at	---	---	514.4	-1.39000	1.47
40420_at	serine/threonine kinase 10	STK10	503.9	-1.40922	1.47
234623_x_at	---	---	293.0	-1.95639	1.46
218124_at	all-trans-13,14-dihydroretinol saturase	RetSat	627.4	-1.21233	1.46
216938_x_at	dopamine receptor D2	DRD2	591.0	-1.26438	1.45
224657_at	ERBB receptor feedback inhibitor 1	ERRFI1	683.6	-1.13944	1.44
202922_at	glutamate-cysteine ligase, catalytic subunit	GCLC	291.0	-1.96385	1.44
1558028_x_at	Translocation associated membrane protein 1	TRAM1	596.5	-1.25631	1.43

211599_x_at	met proto-oncogene (hepatocyte growth factor receptor) /// met proto-oncogene (hepatocyte growth factor receptor)	MET	707.4	-1.11110	1.42
202668_at	ephrin-B2	EFNB2	702.5	-1.11684	1.42
220755_s_at	chromosome 6 open reading frame 48	C6orf48	614.7	-1.23002	1.41
201919_at	Solute carrier family 25, member 36	FLJ10618	440.3	-1.53776	1.41
217996_at	pleckstrin homology-like domain, family A, member 1	PHLDA1	311.9	-1.88953	1.40
223297_at	hypothetical protein MGC4268	MGC4268	425.4	-1.57146	1.40
210513_s_at	vascular endothelial growth factor	VEGF	302.4	-1.92239	1.40
225102_at	monoglyceride lipase	MGLL	310.3	-1.89477	1.39
211653_x_at	aldo-keto reductase family 1, member C2 (dihydrodiol dehydrogenase 2; bile acid binding protein; 3-alpha hydroxysteroid dehydrogenase, type III) /// aldo-keto reductase family 1, member C2 (dihydrodiol dehydrogenase 2; bile acid binding protein; 3-alpha h	AKR1C2	362.9	-1.73106	1.38
232406_at	Jagged 1 (Alagille syndrome)	JAG1	541.5	-1.34281	1.38
1554132_a_at	KIAA1128	KIAA1128	679.1	-1.14500	1.37
226572_at	Suppressor of cytokine signaling 7	SOCS7	387.9	-1.66324	1.35
1564031_a_at	chromosome 5 open reading frame 16	C5orf16	501.5	-1.41356	1.34
1552575_a_at	chromosome 6 open reading frame 141	C6orf141	530.6	-1.36141	1.30
205623_at	aldehyde dehydrogenase 3 family, member A1	ALDH3A1	293.8	-1.95361	1.30
201057_s_at	golgi autoantigen, golgin subfamily b, macrogolgin (with transmembrane signal), 1	GOLGB1	427.5	-1.56655	1.30
203506_s_at	mediator of RNA polymerase II transcription, subunit 12 homolog (yeast)	MED12	345.4	-1.78220	1.29
204272_at	lectin, galactoside-binding, soluble, 4 (galectin 4)	LGALS4	566.6	-1.30183	1.27
202241_at	tribbles homolog 1 (Drosophila)	TRIB1	409.5	-1.60897	1.25
208708_x_at	eukaryotic translation initiation factor 5	EIF5	490.5	-1.43442	1.24
221957_at	Pyruvate dehydrogenase kinase, isoenzyme 3	PDK3	429.7	-1.56152	1.22
200831_s_at	stearoyl-CoA desaturase (delta-9-desaturase)	SCD	370.9	-1.70882	1.22
224558_s_at	metastasis associated lung adenocarcinoma transcript 1 (non-coding RNA)	MALAT1	307.8	-1.90352	1.22
201926_s_at	decay accelerating factor for complement (CD55, Cromer blood group system)	DAF	290.6	-1.96528	1.21
212878_s_at	kinesin 2	KNS2	404.0	-1.62260	1.20
218368_s_at	tumor necrosis factor receptor superfamily, member 12A	TNFRSF12A	306.0	-1.90972	1.20
212729_at	discs, large homolog 3 (neuroendocrine-dlg, Drosophila)	DLG3	370.5	-1.70981	1.19

213428_s_at	collagen, type VI, alpha 1	COL6A1	325.6	-1.84386	1.19
227372_s_at	BAI1-associated protein 2-like 1	BAIAP2L1	302.2	-1.92309	1.18
227224_at	Ral GEF with PH domain and SH3 binding motif 2	RALGPS2	444.1	-1.52929	1.18
202123_s_at	v-abl Abelson murine leukemia viral oncogene homolog 1	ABL1	284.2	-1.98945	1.18
1555167_s_at	pre-B-cell colony enhancing factor 1	PBEF1	400.3	-1.63171	1.16
213656_s_at	kinesin 2	KNS2	356.3	-1.74999	1.12
227616_at	B-cell CLL/lymphoma 9-like	BCL9L	308.4	-1.90153	1.12
219045_at	ras homolog gene family, member F (in filopodia)	RHOF	338.4	-1.80352	1.12
213807_x_at	met proto-oncogene (hepatocyte growth factor receptor)	MET	479.1	-1.45666	1.11
209386_at	transmembrane 4 L six family member 1	TM4SF1	283.1	-1.99381	1.09
211535_s_at	fibroblast growth factor receptor 1 (fms-related tyrosine kinase 2, Pfeiffer syndrome)	FGFR1	286.4	-1.98119	1.08
212942_s_at	KIAA1199	KIAA1199	310.8	-1.89314	1.08
227027_at	Clone 114 tumor rejection antigen	---	345.8	-1.78078	1.08
201908_at	dishevelled, dsh homolog 3 (Drosophila)	DVL3	335.4	-1.81265	1.08
212689_s_at	jumonji domain containing 1A	JMJD1A	361.6	-1.73489	1.06
241355_at	hairless homolog (mouse)	HR	321.1	-1.85871	1.01
217496_s_at	insulin-degrading enzyme	IDE	310.3	-1.89485	1.01
218070_s_at	GDP-mannose pyrophosphorylase A	GMPPA	298.6	-1.93589	0.99
217826_s_at	ubiquitin-conjugating enzyme E2, J1 (UBC6 homolog, yeast)	UBE2J1	301.5	-1.92570	0.96
224831_at	cytoplasmic polyadenylation element binding protein 4	CPEB4	410.6	-1.60637	0.86
217862_at	protein inhibitor of activated STAT, 1	PIAS1	319.2	-1.86494	-0.80
205078_at	phosphatidylinositol glycan, class F	PIGF	304.6	-1.91466	-0.81
228722_at	HMT1 hnRNP methyltransferase-like 1 (S. cerevisiae)	HRMT1L1	282.5	-1.99580	-0.88
221829_s_at	transportin 1	TNPO1	296.4	-1.94390	-0.90
208615_s_at	protein tyrosine phosphatase type IVA, member 2	PTP4A2	293.5	-1.95444	-0.90
202377_at	leptin receptor /// leptin receptor overlapping transcript	LEPR /// LEPROT	356.0	-1.75079	-0.90
204957_at	origin recognition complex, subunit 5-like (yeast)	ORC5L	286.2	-1.98175	-0.92
209064_x_at	poly(A) binding protein interacting protein 1	PAIP1	295.9	-1.94563	-0.92
210461_s_at	actin binding LIM protein 1	ABLIM1	300.2	-1.93035	-0.93
218929_at	collaborates/cooperates with ARF (alternate reading frame) protein	CARF	307.2	-1.90560	-0.94
227514_at	Hypothetical protein LOC162073	LOC162073	287.2	-1.97821	-0.94
220060_s_at	hypothetical protein FLJ20641	FLJ20641	329.4	-1.83155	-0.94
225830_at	PDZ domain containing 8	PDZK8	305.2	-1.91247	-0.95

1557352_at	Squalene epoxidase	SQLE	336.7	-1.80864	-0.95
208638_at	protein disulfide isomerase family A, member 6	PDIA6	291.8	-1.96085	-0.96
204391_x_at	tripartite motif-containing 24	TRIM24	289.5	-1.96932	-0.97
218721_s_at	chromosome 1 open reading frame 27	C1orf27	283.0	-1.99418	-0.97
229394_s_at	Glucocorticoid receptor DNA binding factor 1	GRLF1	306.5	-1.90796	-0.97
211697_x_at	putative 28 kDa protein /// putative 28 kDa protein	LOC56902	321.2	-1.85830	-0.98
201762_s_at	proteasome (prosome, macropain) activator subunit 2 (PA28 beta)	PSME2	313.5	-1.88406	-0.98
206989_s_at	splicing factor, arginine/serine-rich 2, interacting protein	SFRS2IP	287.7	-1.97619	-0.98
201407_s_at	protein phosphatase 1, catalytic subunit, beta isoform	PPP1CB	391.5	-1.65409	-0.98
239034_at	chromosome X open reading frame 24	CXorf24	287.5	-1.97705	-0.98
213024_at	TATA element modulatory factor 1	TMF1	375.5	-1.69618	-0.98
235907_at	Transcribed locus	---	386.0	-1.66823	-0.98
225419_at	chromosome 7 open reading frame 11	C7orf11	343.6	-1.78742	-0.99
223288_at	ubiquitin specific peptidase 38	USP38	329.9	-1.83001	-0.99
208249_s_at	TDP-glucose 4,6-dehydratase	TGDS	290.3	-1.96644	-0.99
218118_s_at	translocase of inner mitochondrial membrane 23 homolog (yeast)	TIMM23	379.8	-1.68471	-1.00
205091_x_at	RecQ protein-like (DNA helicase Q1-like)	RECQL	334.9	-1.81419	-1.00
214315_x_at	calreticulin	CALR	292.8	-1.95723	-1.00
228991_at	Cell division cycle 2-like 5 (cholinesterase-related cell division controller)	CDC2L5	321.7	-1.85673	-1.00
236109_at	RNA pseudouridylate synthase domain containing 4	FLJ14494	304.1	-1.91661	-1.00
230490_x_at	Ras suppressor protein 1	RSU1	329.8	-1.83027	-1.01
210054_at	chromosome 4 open reading frame 15	C4orf15	301.2	-1.92658	-1.01
218662_s_at	chromosome condensation protein G	HCAP-G	293.8	-1.95362	-1.02
1552921_a_at	fidgetin-like 1	FIGNL1	397.3	-1.63913	-1.02
212290_at	solute carrier family 7 (cationic amino acid transporter, y+ system), member 1	SLC7A1	315.5	-1.87709	-1.03
212724_at	Rho family GTPase 3	RND3	294.3	-1.95178	-1.03
224173_s_at	mitochondrial ribosomal protein L30	MRPL30	415.8	-1.59395	-1.03
214642_x_at	melanoma antigen family A, 5	MAGEA5	300.9	-1.92769	-1.03
226635_at	Hypothetical gene supported by AK091718	---	293.0	-1.95656	-1.04
235006_at	similar to RIKEN cDNA A430101B06 gene	MGC13017	432.9	-1.55419	-1.04
222472_at	aftiphilin protein	AFTIPHILIN	306.2	-1.90902	-1.04
212824_at	far upstream element (FUSE) binding protein 3	FUBP3	306.2	-1.90914	-1.04
201577_at	non-metastatic cells 1, protein (NM23A) expressed in	NME1	297.5	-1.93995	-1.04

233559_s_at	WD repeat and FYVE domain containing 1	WDFY1	286.3	-1.98143	-1.04
217427_s_at	HIR histone cell cycle regulation defective homolog A ( <i>S. cerevisiae</i> )	HIRA	291.1	-1.96357	-1.05
208913_at	golgi associated, gamma adaptin ear containing, ARF binding protein 2	GGA2	287.9	-1.97559	-1.05
217743_s_at	transmembrane protein 30A	TMEM30A	355.6	-1.75189	-1.05
225556_at	hypothetical protein LOC203547	LOC203547	292.4	-1.95863	-1.05
200026_at	ribosomal protein L34 /// ribosomal protein L34 /// similar to ribosomal protein L34; 60S ribosomal protein L34 /// similar to ribosomal protein L34; 60S ribosomal protein L34	RPL34 /// LOC342994	284.5	-1.98824	-1.05
202678_at	general transcription factor IIA, 2, 12kDa	GTF2A2	328.7	-1.83381	-1.05
222250_s_at	chromosome 1 open reading frame 73	C1orf73	317.6	-1.87013	-1.05
225092_at	rabaptin, RAB GTPase binding effector protein 1	RABEP1	317.3	-1.87113	-1.06
219123_at	zinc finger protein 232	ZNF232	312.6	-1.88710	-1.06
209945_s_at	glycogen synthase kinase 3 beta	GSK3B	322.8	-1.85308	-1.06
212082_s_at	myosin, light polypeptide 6, alkali, smooth muscle and non-muscle	MYL6	285.0	-1.98651	-1.06
228454_at	ligand-dependent corepressor	MLR2	292.2	-1.95927	-1.06
202372_at	Isoleucine-tRNA synthetase 2, mitochondrial	FLJ10326	342.4	-1.79107	-1.06
212410_at	EF-hand domain family, member A1	EFHA1	335.6	-1.81224	-1.06
219933_at	glutaredoxin 2	GLRX2	321.1	-1.85861	-1.06
201151_s_at	muscleblind-like ( <i>Drosophila</i> )	MBNL1	329.5	-1.83135	-1.06
236641_at	kinesin family member 14	KIF14	302.5	-1.92205	-1.06
225101_s_at	sorting nexin 14	SNX14	364.2	-1.72726	-1.06
203367_at	dual specificity phosphatase 14	DUSP14	355.1	-1.75357	-1.06
208739_x_at	SMT3 suppressor of mif two 3 homolog 2 (yeast)	SUMO2	310.6	-1.89372	-1.07
202464_s_at	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3	PFKFB3	302.6	-1.92175	-1.07
229490_s_at	IQ motif containing GTPase activating protein 3	IQGAP3	372.1	-1.70537	-1.07
203455_s_at	spermidine/spermine N1-acetyltransferase	SAT	297.1	-1.94163	-1.07
236241_at	Mediator of RNA polymerase II transcription, subunit 31 homolog (yeast)	MED31	331.9	-1.82390	-1.07
205990_s_at	wingless-type MMTV integration site family, member 5A	WNT5A	463.6	-1.48793	-1.07
225099_at	F-box protein 45	FBXO45	311.9	-1.88944	-1.08
202514_at	discs, large homolog 1 ( <i>Drosophila</i> )	DLG1	301.8	-1.92477	-1.08
208876_s_at	p21 (CDKN1A)-activated kinase 2	PAK2	339.3	-1.80061	-1.08
218354_at	hematopoietic stem/progenitor cells 176	HSPC176	281.7	-1.99899	-1.08
201310_s_at	chromosome 5 open reading frame 13	C5orf13	368.7	-1.71470	-1.08
212330_at	transcription factor Dp-1	TFDP1	282.8	-1.99479	-1.08

207573_x_at	ATP synthase, H <sup>+</sup> transporting, mitochondrial F0 complex, subunit g	ATP5L	284.8	-1.98705	-1.08
213457_at	malignant fibrous histiocytoma amplified sequence 1	MFHAS1	281.8	-1.99857	-1.08
213016_at	Bobby sox homolog (Drosophila)	BBX	300.4	-1.92955	-1.08
37965_at	parvin, beta	PARVB	292.0	-1.96007	-1.08
221381_s_at	mortality factor 4 like 1 /// mortality factor 4	MORF4L1 /// MORF4	340.9	-1.79590	-1.09
40284_at	forkhead box A2	FOXA2	331.4	-1.82530	-1.09
219329_s_at	chromosome 2 open reading frame 28	C2orf28	302.5	-1.92222	-1.09
212279_at	hypothetical protein MAC30	MAC30	293.0	-1.95655	-1.09
226794_at	syntaxin binding protein 5 (tomosyn)	STXBP5	327.7	-1.83717	-1.09
221263_s_at	splicing factor 3b, subunit 5, 10kDa /// splicing factor 3b, subunit 5, 10kDa	SF3B5	289.2	-1.97036	-1.09
227373_at	hypothetical protein LOC146517	LOC146517	390.2	-1.65745	-1.09
200032_s_at	ribosomal protein L9 /// ribosomal protein L9	RPL9	363.4	-1.72953	-1.09
203960_s_at	chromosome 1 open reading frame 41	C1orf41	287.2	-1.97821	-1.10
225264_at	arginyl-tRNA synthetase-like	RARSL	337.7	-1.80575	-1.10
201634_s_at	outer mitochondrial membrane cytochrome b5	CYB5-M	397.3	-1.63916	-1.10
202396_at	transcription elongation regulator 1	TCERG1	285.3	-1.98512	-1.10
201658_at	ADP-ribosylation factor-like 1	ARL1	396.9	-1.64011	-1.10
225737_s_at	F-box protein 22	FBXO22	443.1	-1.53162	-1.10
200853_at	H2A histone family, member Z	H2AFZ	286.9	-1.97906	-1.11
214855_s_at	GTPase activating Rap/RanGAP domain-like 1	GARNL1	375.7	-1.69574	-1.11
228562_at	Zinc finger and BTB domain containing 10	ZBTB10	477.8	-1.45922	-1.11
200037_s_at	chromobox homolog 3 (HP1 gamma homolog, Drosophila) /// chromobox homolog 3 (HP1 gamma homolog, Drosophila)	CBX3	326.1	-1.84224	-1.11
239376_at	CDNA clone IMAGE:4333081	---	362.4	-1.73241	-1.11
241885_at	TAF15 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 68kDa	TAF15	394.5	-1.64635	-1.11
223642_at	Zic family member 2 (odd-paired homolog, Drosophila)	ZIC2	389.3	-1.65961	-1.11
203011_at	inositol(myo)-1(or 4)-monophosphatase 1	IMPA1	296.8	-1.94252	-1.11
208289_s_at	etoposide induced 2.4 mRNA	EI24	416.4	-1.59255	-1.11
239487_at	DKFZP564F0522 protein	DKFZP564F0522	302.1	-1.92366	-1.12
201433_s_at	phosphatidylserine synthase 1	PTDSS1	363.7	-1.72891	-1.12
212638_s_at	WW domain containing E3 ubiquitin protein ligase 1	WWP1	417.8	-1.58924	-1.12
219017_at	ethanolamine kinase 1	ETNK1	357.4	-1.74677	-1.12

223595_at	AD031 protein	AD031	327.9	-1.83636	-1.12
219174_at	coiled-coil domain containing 2	CCDC2	319.6	-1.86342	-1.12
226826_at	LSM11, U7 small nuclear RNA associated	LSM11	285.2	-1.98555	-1.12
203351_s_at	origin recognition complex, subunit 4-like (yeast)	ORC4L	350.1	-1.76814	-1.12
225625_at	similar to hypothetical protein 9530023G02	MGC90512	454.6	-1.50671	-1.12
212333_at	DKFZP564F0522 protein	DKFZP564F0522	330.9	-1.82701	-1.12
213649_at	splicing factor, arginine/serine-rich 7, 35kDa	SFRS7	305.7	-1.91091	-1.12
217814_at	GK001 protein	GK001	301.6	-1.92538	-1.12
226452_at	pyruvate dehydrogenase kinase, isoenzyme 1	PDK1	411.1	-1.60506	-1.12
227621_at	Wilms tumor 1 associated protein	WTAP	322.3	-1.85451	-1.12
225971_at	DDHD domain containing 1	DDHD1	292.6	-1.95805	-1.12
209103_s_at	ubiquitin fusion degradation 1 like (yeast)	UFD1L	311.1	-1.89210	-1.12
202345_s_at	fatty acid binding protein 5 (psoriasis-associated)	FABP5	398.6	-1.63584	-1.12
225986_x_at	cleavage and polyadenylation specific factor 2, 100kDa	CPSF2	295.4	-1.94765	-1.12
212329_at	SREBP cleavage-activating protein	SCAP	292.3	-1.95907	-1.12
209020_at	chromosome 20 open reading frame 111	C20orf111	343.3	-1.78858	-1.12
238156_at	Ribosomal protein S6	RPS6	318.0	-1.86901	-1.13
64900_at	carbohydrate (N-acetylglucosamine 6-O) sulfotransferase 5 /// hypothetical protein MGC15429	CHST5 /// MGC15429	339.4	-1.80052	-1.13
208841_s_at	Ras-GTPase activating protein SH3 domain-binding protein 2	G3BP2	343.1	-1.78918	-1.13
222992_s_at	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 9, 22kDa	NDUFB9	366.0	-1.72233	-1.13
200887_s_at	signal transducer and activator of transcription 1, 91kDa	STAT1	351.7	-1.76343	-1.13
219015_s_at	glycosyltransferase 28 domain containing 1	GLT28D1	423.8	-1.57498	-1.13
218319_at	pellino homolog 1 (Drosophila)	PELI1	314.2	-1.88152	-1.14
235343_at	Hypothetical protein FLJ12505	FLJ12505	307.9	-1.90318	-1.14
226399_at	DnaJ (Hsp40) homolog, subfamily B, member 14	FLJ14281	409.7	-1.60865	-1.14
202905_x_at	nibrin	NBN	388.2	-1.66249	-1.14
203428_s_at	ASF1 anti-silencing function 1 homolog A (S. cerevisiae)	ASF1A	420.1	-1.58367	-1.14
223113_at	hypothetical protein HSPC196	HSPC196	354.6	-1.75503	-1.14
203607_at	inositol polyphosphate-5-phosphatase F disrupter of silencing 10	INPP5F	368.4	-1.71578	-1.14
209486_at	disrupter of silencing 10	SAS10	355.5	-1.75236	-1.14
216438_s_at	thymosin, beta 4, X-linked /// thymosin-like 3	TMSB4X /// TMSL3	326.5	-1.84097	-1.14
218859_s_at	chromosome 20 open reading frame 6	C20orf6	354.6	-1.75486	-1.14

200080_s_at	H3 histone, family 3A /// H3 histone, family 3A /// H3 histone, family 3A pseudogene /// H3 histone, family 3A pseudogene	H3F3A /// LOC440926	323.4	-1.85108	-1.14
201821_s_at	translocase of inner mitochondrial membrane 17 homolog A (yeast)	TIMM17A	353.8	-1.75711	-1.15
214590_s_at	ubiquitin-conjugating enzyme E2D 1 (UBC4/5 homolog, yeast)	UBE2D1	289.4	-1.96962	-1.15
226631_at	similar to CG9643-PA	LOC399818	396.7	-1.64078	-1.15
232271_at	hepatocyte nuclear factor 4, gamma	HNF4G	437.4	-1.54416	-1.15
233873_x_at	PAP associated domain containing 1	PAPD1	289.6	-1.96918	-1.15
201999_s_at	t-complex-associated-testis-expressed 1-like 1	TCTEL1	308.7	-1.90049	-1.15
208296_x_at	tumor necrosis factor, alpha-induced protein 8	TNFAIP8	402.9	-1.62510	-1.15
1552664_at	folliculin	FLCN	410.9	-1.60553	-1.15
204354_at	POT1 protection of telomeres 1 homolog (S. pombe)	POT1	361.6	-1.73484	-1.15
225903_at	CDC91 cell division cycle 91-like 1 (S. cerevisiae)	CDC91L1	347.1	-1.77705	-1.15
218687_s_at	mucin 13, epithelial transmembrane	MUC13	409.3	-1.60959	-1.15
1555594_a_at	muscleblind-like (Drosophila)	MBNL1	446.2	-1.52484	-1.16
225563_at	PABP1-dependent poly A-specific ribonuclease subunit PAN3	PAN3	418.0	-1.58878	-1.16
224569_s_at	interferon regulatory factor 2 binding protein 2	IRF2BP2	352.5	-1.76096	-1.16
226520_at	Transcribed locus	---	366.2	-1.72190	-1.16
218700_s_at	RAB7, member RAS oncogene family-like 1	RAB7L1	327.7	-1.83703	-1.16
236259_at	serine/threonine kinase 4	STK4	326.0	-1.84273	-1.16
213446_s_at	IQ motif containing GTPase activating protein 1	IQGAP1	365.5	-1.72369	-1.16
39729_at	peroxiredoxin 2	PRDX2	284.7	-1.98768	-1.16
208074_s_at	adaptor-related protein complex 2, sigma 1 subunit	AP2S1	426.7	-1.56833	-1.16
235253_at	RAD1 homolog (S. pombe)	RAD1	380.9	-1.68169	-1.16
223892_s_at	transmembrane BAX inhibitor motif containing 4	TMBIM4	291.6	-1.96172	-1.16
222754_at	tRNA nucleotidyl transferase, CCA-adding, 1	TRNT1	323.8	-1.84988	-1.16
225935_at	CDNA clone IMAGE:4769453	---	371.8	-1.70617	-1.17
226917_s_at	anaphase promoting complex subunit 4	ANAPC4	319.4	-1.86435	-1.17
218519_at	solute carrier family 35, member A5	SLC35A5	463.1	-1.48889	-1.17
218203_at	asparagine-linked glycosylation 5 homolog (yeast, dolichyl-phosphate beta-glucosyltransferase)	ALG5	376.8	-1.69256	-1.17
64408_s_at	calmodulin-like 4	CALML4	352.9	-1.75988	-1.17

200699_at	KDEL (Lys-Asp-Glu-Leu) endoplasmic reticulum protein retention receptor 2	KDELR2	408.5	-1.61158	-1.17
208805_at	proteasome (prosome, macropain) subunit, alpha type, 6	PSMA6	309.9	-1.89644	-1.17
227942_s_at	postsynaptic protein CRIPT	CRIP1	402.1	-1.62732	-1.17
229520_s_at	chromosome 14 open reading frame 118	C14orf118	349.8	-1.76888	-1.17
209753_s_at	thymopoietin	TMPO	433.7	-1.55250	-1.17
221791_s_at	hypothetical protein HSPC016	HSPC016	289.0	-1.97146	-1.17
220934_s_at	hypothetical protein MGC3196	MGC3196	415.9	-1.59357	-1.17
218633_x_at	abhydrolase domain containing 10	ABHD10	291.2	-1.96323	-1.18
203613_s_at	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 6, 17kDa	NDUFB6	295.0	-1.94927	-1.18
1561757_a_at	hypothetical protein LOC283352	LOC283352	286.4	-1.98118	-1.18
214433_s_at	selenium binding protein 1 /// selenium binding protein 1	SELENBP1	282.6	-1.99575	-1.18
227551_at	chromosome 9 open reading frame 77	C9orf77	437.8	-1.54319	-1.18
213507_s_at	karyopherin (importin) beta 1	KPNB1	409.6	-1.60868	-1.18
225284_at	hypothetical protein LOC144871	LOC144871	396.5	-1.64125	-1.18
201524_x_at	ubiquitin-conjugating enzyme E2N (UBC13 homolog, yeast)	UBE2N	284.7	-1.98757	-1.18
226120_at	tetratricopeptide repeat domain 8	TTC8	289.1	-1.97100	-1.18
224511_s_at	thioredoxin-like 5 /// thioredoxin-like 5	TXNL5	346.6	-1.77864	-1.18
204186_s_at	peptidylprolyl isomerase D (cyclophilin D)	PPID	295.2	-1.94850	-1.18
225313_at	chromosome 20 open reading frame 177	C20orf177	390.6	-1.65625	-1.18
200839_s_at	cathepsin B	CTSB	342.7	-1.79040	-1.18
201019_s_at	eukaryotic translation initiation factor 1A, X-linked	EIF1AX	287.2	-1.97794	-1.18
219979_s_at	hypothetical protein HSPC138	HSPC138	325.4	-1.84466	-1.18
224523_s_at	hypothetical protein MGC4308 /// hypothetical protein MGC4308	MGC4308	361.2	-1.73602	-1.19
217761_at	membrane-type 1 matrix metalloproteinase cytoplasmic tail binding protein-1	MTCBP-1	297.4	-1.94037	-1.19
225427_s_at	apolipoprotein A-I binding protein	APOA1BP	341.6	-1.79377	-1.19
221830_at	RAP2A, member of RAS oncogene family	RAP2A	427.3	-1.56700	-1.19
203067_at	pyruvate dehydrogenase complex, component X	PDHX	371.0	-1.70851	-1.19
226385_s_at	chromosome 7 open reading frame 30	C7orf30	301.7	-1.92491	-1.19
225367_at	phosphoglucomutase 2	PGM2	445.2	-1.52700	-1.19
224935_at	Eukaryotic translation initiation factor 2, subunit 3 gamma, 52kDa	EIF2S3	339.2	-1.80110	-1.19
226481_at	Vpr-binding protein	VprBP	394.5	-1.64623	-1.19
200772_x_at	prothymosin, alpha (gene sequence 28)	PTMA	398.7	-1.63566	-1.19
205799_s_at	solute carrier family 3 (cystine, dibasic and neutral amino acid transporters, activator of cystine, dibasic and neutral amino acid transport), member 1	SLC3A1	485.5	-1.44409	-1.19

201091_s_at	chromobox homolog 3 (HP1 gamma homolog, Drosophila)	CBX3	399.5	-1.63380	-1.19
201906_s_at	CTD (carboxy-terminal domain, RNA polymerase II, polypeptide A) small phosphatase-like	CTDSPL	307.4	-1.90478	-1.19
218170_at	isochorismatase domain containing 1	ISOC1	324.5	-1.84736	-1.19
202271_at	F-box protein 28	FBXO28	288.5	-1.97305	-1.19
224751_at	Full-length cDNA clone CS0DB005YG10 of Neuroblastoma Cot 10-normalized of Homo sapiens (human)	---	457.8	-1.50010	-1.20
233241_at	chromosome 20 open reading frame 19	C20orf19	307.3	-1.90511	-1.20
239392_s_at	Pogo transposable element with KRAB domain	POGK	447.8	-1.52131	-1.20
200818_at	ATP synthase, H <sup>+</sup> transporting, mitochondrial F1 complex, O subunit (oligomycin sensitivity conferring protein)	ATP5O	399.5	-1.63365	-1.20
225413_at	upregulated during skeletal muscle growth 5	USMG5	415.8	-1.59381	-1.20
231866_at	leucyl/cystinyl aminopeptidase	LNPEP	422.4	-1.57843	-1.20
227369_at	SERPINE1 mRNA binding protein 1	SERBP1	422.2	-1.57891	-1.20
225716_at	Full-length cDNA clone CS0DK008YI09 of HeLa cells Cot 25-normalized of Homo sapiens (human)	---	363.0	-1.73078	-1.20
225910_at	hypothetical protein LOC284019	LOC284019	330.3	-1.82892	-1.20
232899_at	ribosomal protein L23a pseudogene 7 /// family with sequence similarity 41, member C /// similar to RPL23AP7 protein	RPL23AP7 /// FAM41C /// MGC70863	298.6	-1.93618	-1.20
225334_at	chromosome 10 open reading frame 32	C10orf32	382.8	-1.67656	-1.20
212458_at	sprouty-related, EVH1 domain containing 2	SPRED2	324.9	-1.84613	-1.20
228338_at	hypothetical protein LOC120376	LOC120376	317.4	-1.87076	-1.20
218826_at	solute carrier family 35, member F2	SLC35F2	284.4	-1.98870	-1.20
203621_at	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 5, 16kDa	NDUFB5	378.4	-1.68839	-1.21
218040_at	PRP38 pre-mRNA processing factor 38 (yeast) domain containing B	PRPF38B	349.8	-1.76886	-1.21
244828_x_at	hypothetical protein BC008207	LOC92345	380.7	-1.68220	-1.21
224938_at	82-kD FMRP Interacting Protein	182-FIP	479.4	-1.45607	-1.21
208648_at	valosin-containing protein	VCP	384.4	-1.67237	-1.21
217992_s_at	EF-hand domain family, member D2	EFHD2	392.9	-1.65033	-1.21
208821_at	small nuclear ribonucleoprotein polypeptides B and B1	SNRPB	373.1	-1.70265	-1.21
219226_at	CDC2-related protein kinase 7	CRK7	362.2	-1.73311	-1.21
203355_s_at	pleckstrin and Sec7 domain containing 3	PSD3	461.0	-1.49339	-1.21
216384_x_at	similar to prothymosin alpha	LOC440085	340.7	-1.79630	-1.21
225127_at	KIAA1423	KIAA1423	401.6	-1.62838	-1.21

226963_at	basic transcription factor 3-like 4	BTF3L4	358.3	-1.74408	-1.21
209714_s_at	cyclin-dependent kinase inhibitor 3 (CDK2-associated dual specificity phosphatase)	CDKN3	416.5	-1.59227	-1.21
223230_at	PRP38 pre-mRNA processing factor 38 (yeast) domain containing A	PRPF38A	363.8	-1.72851	-1.21
226990_at	membrane component, chromosome 11, surface marker 1	M11S1	344.7	-1.78413	-1.22
1557081_at	RNA binding motif protein 25	RBM25	362.9	-1.73103	-1.22
239143_x_at	ring finger protein 138	RNF138	491.7	-1.43218	-1.22
218846_at	cofactor required for Sp1 transcriptional activation, subunit 3, 130kDa	CRSP3	314.6	-1.88017	-1.22
211085_s_at	serine/threonine kinase 4 /// serine/threonine kinase 4	STK4	410.5	-1.60656	-1.22
205048_s_at	phosphoserine phosphatase	PSPH	399.1	-1.63468	-1.22
224687_at	ankyrin repeat and IBR domain containing 1	ANKIB1	293.9	-1.95312	-1.22
210453_x_at	ATP synthase, H <sup>+</sup> transporting, mitochondrial F0 complex, subunit g	ATP5L	307.6	-1.90434	-1.22
211921_x_at	prothymosin, alpha (gene sequence 28) /// prothymosin, alpha (gene sequence 28)	PTMA	410.0	-1.60793	-1.22
202824_s_at	transcription elongation factor B (SIII), polypeptide 1 (15kDa, elongin C)	TCEB1	357.5	-1.74645	-1.22
218480_at	hypothetical protein FLJ21839	FLJ21839	340.9	-1.79589	-1.22
223207_x_at	phosphohistidine phosphatase 1	PHPT1	411.8	-1.60345	-1.22
219062_s_at	zinc finger, CCHC domain containing 2	ZCCHC2	551.4	-1.32645	-1.22
225521_at	anaphase promoting complex subunit 7	ANAPC7	312.1	-1.88861	-1.22
218152_at	high-mobility group 20A	HMG20A	411.7	-1.60380	-1.22
225507_at	chromosome 6 open reading frame 111	C6orf111	336.7	-1.80886	-1.23
224676_at	transmembrane emp24 protein transport domain containing 4	TMED4	302.2	-1.92318	-1.23
217976_s_at	dynein, cytoplasmic, light intermediate polypeptide 1	DNCLI1	339.9	-1.79886	-1.23
214659_x_at	YLP motif containing 1	YLPM1	282.6	-1.99553	-1.23
200099_s_at	ribosomal protein S3A /// ribosomal protein S3A	RPS3A	418.9	-1.58646	-1.23
201065_s_at	general transcription factor II, i /// general transcription factor II, i, pseudogene 1	GTF2I /// GTF2IP1	372.1	-1.70538	-1.23
229322_at	protein phosphatase 2, regulatory subunit B (B56), epsilon isoform	PPP2R5E	353.9	-1.75701	-1.23
205060_at	poly (ADP-ribose) glycohydrolase	PARG	385.2	-1.67041	-1.23
238701_x_at	FLJ45803 protein	FLJ45803	374.6	-1.69860	-1.23
226005_at	Ubiquitin-conjugating enzyme E2G 1 (UBC7 homolog, yeast)	UBE2G1	347.8	-1.77503	-1.23
213677_s_at	PMS1 postmeiotic segregation increased 1 (S. cerevisiae)	PMS1	357.1	-1.74763	-1.23
223114_at	hypothetical protein MGC4767	MGC4767	373.7	-1.70120	-1.23

203852_s_at	survival of motor neuron 1, telomeric /// survival of motor neuron 2, centromeric	SMN1 /// SMN2	451.7	-1.51295	-1.23
223229_at	ubiquitin-conjugating enzyme E2T (putative)	UBE2T	342.6	-1.79068	-1.23
233888_s_at	SLIT-ROBO Rho GTPase activating protein 1	SRGAP1	313.0	-1.88561	-1.23
224715_at	WD repeat domain 34	WDR34	310.2	-1.89532	-1.23
225285_at	branched chain aminotransferase 1, cytosolic	BCAT1	327.9	-1.83658	-1.23
206295_at	interleukin 18 (interferon-gamma-inducing factor)	IL18	395.5	-1.64367	-1.24
223081_at	PHD finger protein 23	PHF23	418.2	-1.58823	-1.24
216246_at	---	---	329.1	-1.83263	-1.24
205167_s_at	cell division cycle 25C	CDC25C	330.2	-1.82922	-1.24
224151_s_at	adenylate kinase 3	AK3	372.8	-1.70349	-1.24
203538_at	calcium modulating ligand	CAMLG	434.3	-1.55109	-1.24
203743_s_at	thymine-DNA glycosylase	TDG	520.8	-1.37863	-1.24
201669_s_at	myristoylated alanine-rich protein kinase C substrate	MARCKS	351.6	-1.76376	-1.24
209563_x_at	calmodulin 1 (phosphorylase kinase, delta)	CALM1	369.4	-1.71296	-1.24
223338_s_at	ATPase inhibitory factor 1	ATPIF1	506.0	-1.40518	-1.24
212547_at	FLJ35348 /// Bromodomain containing 3	FLJ35348 /// BRD3	399.7	-1.63324	-1.24
224780_at	RNA binding motif protein 17	RBM17	405.1	-1.61969	-1.24
219260_s_at	S-phase 2 protein	DERP6	331.1	-1.82635	-1.24
212449_s_at	lysophospholipase I	LYPLA1	283.2	-1.99336	-1.24
223155_at	haloacid dehalogenase-like hydrolase domain containing 2	HDHD2	294.6	-1.95056	-1.24
202158_s_at	CUG triplet repeat, RNA binding protein 2	CUGBP2	443.2	-1.53134	-1.24
218273_s_at	protein phosphatase 2C, magnesium-dependent, catalytic subunit	PPM2C	357.7	-1.74590	-1.24
224972_at	chromosome 20 open reading frame 52	C20orf52	376.1	-1.69457	-1.25
224617_at	---	---	385.9	-1.66855	-1.25
231870_s_at	NMD3 homolog (S. cerevisiae)	CGI-07	456.2	-1.50342	-1.25
221622_s_at	uncharacterized hypothalamus protein HT007	HT007	355.1	-1.75331	-1.25
203745_at	holocytochrome c synthase (cytochrome c heme-lyase)	HCCS	289.5	-1.96932	-1.25
208319_s_at	RNA binding motif (RNP1, RRM) protein 3	RBM3	291.8	-1.96077	-1.25
202356_s_at	general transcription factor IIF, polypeptide 1, 74kDa	GTF2F1	445.1	-1.52717	-1.25
234405_s_at	RNA U, small nuclear RNA export adaptor (phosphorylation regulated)	RNUXA	336.6	-1.80891	-1.25
200694_s_at	DEAD (Asp-Glu-Ala-Asp) box polypeptide 24	DDX24	311.2	-1.89170	-1.25

231812_x_at	RNA U, small nuclear RNA export adaptor (phosphorylation regulated)	RNUXA	287.7	-1.97603	-1.25
219972_s_at	chromosome 14 open reading frame 135	C14orf135	453.8	-1.50857	-1.25
1552628_a_at	hypothetical protein FLJ22313	FLJ22313	398.0	-1.63738	-1.25
209821_at	chromosome 9 open reading frame 26 (NF-HEV)	C9orf26	407.4	-1.61422	-1.25
209610_s_at	solute carrier family 1 (glutamate/neutral amino acid transporter), member 4	SLC1A4	328.3	-1.83513	-1.26
213857_s_at	CD47 antigen (Rh-related antigen, integrin-associated signal transducer)	CD47	415.6	-1.59438	-1.26
220865_s_at	trans-prenyltransferase	TPRT	337.8	-1.80531	-1.26
218357_s_at	translocase of inner mitochondrial membrane 8 homolog B (yeast)	TIMM8B	369.1	-1.71378	-1.26
218080_x_at	Fas (TNFRSF6) associated factor 1	FAF1	334.2	-1.81649	-1.26
201880_at	Ariadne homolog, ubiquitin-conjugating enzyme E2 binding protein, 1 (Drosophila)	ARIH1	354.9	-1.75403	-1.26
224639_at	signal peptide peptidase 3	SPPL3	331.6	-1.82472	-1.26
212239_at	phosphoinositide-3-kinase, regulatory subunit 1 (p85 alpha)	PIK3R1	446.4	-1.52426	-1.26
225761_at	PAP associated domain containing 4	PAPD4	362.5	-1.73222	-1.26
219644_at	NY-REN-58 antigen	NY-REN-58	285.8	-1.98344	-1.27
228273_at	Hypothetical protein FLJ11029	FLJ11029	310.2	-1.89537	-1.27
235381_at	Hepatitis B virus x associated protein	HBXAP	348.2	-1.77378	-1.27
201533_at	catenin (cadherin-associated protein), beta 1, 88kDa	CTNNB1	290.3	-1.96653	-1.27
210757_x_at	disabled homolog 2, mitogen-responsive phosphoprotein (Drosophila)	DAB2	441.2	-1.53567	-1.27
228297_at	Calponin 3, acidic	CNN3	385.2	-1.67030	-1.27
229349_at	lin-28 homolog B (C. elegans)	LIN28B	307.1	-1.90603	-1.27
209317_at	polymerase (RNA) I polypeptide C, 30kDa	POLR1C	334.8	-1.81462	-1.27
209257_s_at	chondroitin sulfate proteoglycan 6 (bamacan)	CSPG6	345.8	-1.78102	-1.27
225734_at	F-box protein 22	FBXO22	481.3	-1.45234	-1.27
41858_at	FGF receptor activating protein 1	FRAG1	441.6	-1.53490	-1.27
201135_at	enoyl Coenzyme A hydratase, short chain, 1, mitochondrial	ECHS1	408.4	-1.61159	-1.27
208667_s_at	suppression of tumorigenicity 13 (colon carcinoma) (Hsp70 interacting protein)	ST13	347.2	-1.77668	-1.27
1554628_at	hypothetical protein LOC126295	LOC126295	486.2	-1.44268	-1.27
234950_s_at	ring finger and WD repeat domain 2	RFWD2	480.7	-1.45349	-1.27
222482_at	single stranded DNA binding protein 3	SSBP3	420.8	-1.58216	-1.27
202028_s_at	---	---	401.5	-1.62867	-1.27
235142_at	zinc finger and BTB domain containing 8	ZBTB8	419.7	-1.58461	-1.27
222578_s_at	ubiquitin-activating enzyme E1-domain containing 1	UBE1DC1	307.6	-1.90412	-1.28

225841_at	chromosome 1 open reading frame 59	C1orf59	401.5	-1.62878	-1.28
212773_s_at	translocase of outer mitochondrial membrane 20 homolog (yeast)	TOMM20	433.6	-1.55257	-1.28
203755_at	BUB1 budding uninhibited by benzimidazoles 1 homolog beta (yeast)	BUB1B	298.3	-1.93696	-1.28
202001_s_at	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 6, 14kDa	NDUFA6	334.4	-1.81594	-1.28
209238_at	syntaxin 3A	STX3A	464.3	-1.48643	-1.28
223631_s_at	chromosome 19 open reading frame 33	C19orf33	360.0	-1.73931	-1.28
222395_s_at	hypothetical protein FLJ13855	FLJ13855	468.5	-1.47779	-1.28
203360_s_at	c-myc binding protein	MYCBP	668.1	-1.15874	-1.28
223215_s_at	chromosome 14 open reading frame 100	C14orf100	394.7	-1.64567	-1.28
226400_at	Cell division cycle 42 (GTP binding protein, 25kDa)	CDC42	424.9	-1.57266	-1.28
203895_at	phospholipase C, beta 4	PLCB4	395.3	-1.64428	-1.28
216095_x_at	myotubularin related protein 1	MTMR1	334.6	-1.81539	-1.29
216211_at	Chromosome 10 open reading frame 18	C10orf18	342.5	-1.79077	-1.29
224691_at	U2AF homology motif (UHM) kinase 1	UHMK1	452.5	-1.51123	-1.29
222685_at	family with sequence similarity 29, member A	FAM29A	290.7	-1.96486	-1.29
212772_s_at	ATP-binding cassette, sub-family A (ABC1), member 2	ABCA2	556.2	-1.31857	-1.29
205541_s_at	G1 to S phase transition 2 /// G1 to S phase transition 2	GSPT2	367.6	-1.71787	-1.29
227447_at	superkiller viralicidic activity 2-like 2 (S. cerevisiae)	SKIV2L2	520.7	-1.37864	-1.29
212076_at	myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila)	MLL	297.8	-1.93890	-1.29
204170_s_at	CDC28 protein kinase regulatory subunit 2	CKS2	509.3	-1.39920	-1.29
236841_at	CXYorf1-related protein	FLJ25222	375.3	-1.69673	-1.29
226075_at	splA/ryanodine receptor domain and SOCS box containing 1	SPSB1	403.2	-1.62448	-1.29
201327_s_at	chaperonin containing TCP1, subunit 6A (zeta 1)	CCT6A	316.2	-1.87475	-1.29
227068_at	phosphoglycerate kinase 1	PGK1	291.9	-1.96057	-1.29
201841_s_at	heat shock 27kDa protein 1	HSPB1	318.2	-1.86833	-1.29
200816_s_at	platelet-activating factor acetylhydrolase, isoform Ib, alpha subunit 45kDa	PAFAH1B1	363.8	-1.72864	-1.30
235067_at	muskelin 1, intracellular mediator containing kelch motifs	MKLN1	431.4	-1.55765	-1.30
213051_at	zinc finger CCCH-type, antiviral 1	ZC3HAV1	294.2	-1.95196	-1.30
230906_at	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 10 (GalNAc-T10)	GALNT10	311.7	-1.89005	-1.30
202133_at	WW domain containing transcription regulator 1	WWTR1	463.7	-1.48770	-1.30

208925_at	chromosome 3 open reading frame 4	C3orf4	346.4	-1.77906	-1.30
1553528_a_at	TAF5 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 100kDa	TAF5	462.3	-1.49062	-1.30
224330_s_at	mitochondrial ribosomal protein L27 /// mitochondrial ribosomal protein L27	MRPL27	446.0	-1.52515	-1.30
213616_at	Chromosome 18 open reading frame 10	C18orf10	334.8	-1.81457	-1.30
209657_s_at	heat shock transcription factor 2	HSF2	374.5	-1.69884	-1.30
226241_s_at	mitochondrial ribosomal protein L52	MRPL52	461.1	-1.49317	-1.30
208745_at	ATP synthase, H <sup>+</sup> transporting, mitochondrial F0 complex, subunit g	ATP5L	437.8	-1.54328	-1.30
223038_s_at	family with sequence similarity 60, member A	FAM60A	345.0	-1.78340	-1.30
242693_at	Cell division cycle 2-like 5 (cholinesterase-related cell division controller)	CDC2L5	387.0	-1.66568	-1.30
209041_s_at	ubiquitin-conjugating enzyme E2G 2 (UBC7 homolog, yeast)	UBE2G2	349.8	-1.76904	-1.30
201695_s_at	nucleoside phosphorylase	NP	477.4	-1.46002	-1.31
224731_at	high-mobility group box 1	HMGB1	362.8	-1.73143	-1.31
225009_at	chemokine-like factor superfamily 4	CKLFSF4	459.7	-1.49596	-1.31
233092_s_at	DKFZP434B061 protein	DKFZP434B061	366.9	-1.71980	-1.31
208728_s_at	cell division cycle 42 (GTP binding protein, 25kDa)	CDC42	446.4	-1.52432	-1.31
208847_s_at	alcohol dehydrogenase 5 (class III), chi polypeptide	ADH5	394.2	-1.64701	-1.31
226767_s_at	fumarylacetoacetate hydrolase domain containing 1	FAHD1	443.0	-1.53167	-1.31
201297_s_at	MOB1, Mps One Binder kinase activator-like 1B (yeast)	MOBK1B	362.6	-1.73184	-1.31
232541_at	Epidermal growth factor receptor (erythroblastic leukemia viral (v-erb-b) oncogene homolog, avian)	EGFR	390.0	-1.65793	-1.31
224914_s_at	cytokine induced protein 29 kDa	CIP29	348.4	-1.77301	-1.31
210111_s_at	KIAA0265 protein	KIAA0265	300.8	-1.92826	-1.31
218889_at	nucleolar complex associated 3 homolog (S. cerevisiae)	NOC3L	364.4	-1.72676	-1.31
212808_at	nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 2 interacting protein	NFATC2IP	384.4	-1.67231	-1.32
57739_at	dead end homolog 1 (zebrafish)	DND1	314.1	-1.88198	-1.32
219350_s_at	diablo homolog (Drosophila)	DIABLO	428.6	-1.56395	-1.32
212451_at	KIAA0256 gene product	KIAA0256	311.0	-1.89260	-1.32
238021_s_at	hypothetical gene supported by AF275804	LOC388279	498.2	-1.41977	-1.32
203820_s_at	IGF-II mRNA-binding protein 3	IMP-3	491.9	-1.43178	-1.32
204634_at	NIMA (never in mitosis gene a)-related kinase 4	NEK4	536.9	-1.35064	-1.32

202595_s_at	leptin receptor overlapping transcript-like 1	LEPROTL1	466.2	-1.48258	-1.32
218708_at	NTF2-like export factor 1	NXT1	439.0	-1.54054	-1.32
215905_s_at	WD repeat domain 57 (U5 snRNP specific)	WDR57	317.1	-1.87185	-1.32
206613_s_at	TATA box binding protein (TBP)-associated factor, RNA polymerase I, A, 48kDa	TAF1A	520.5	-1.37909	-1.32
209944_at	zinc finger protein 410	ZNF410	312.9	-1.88587	-1.32
212765_at	calmodulin regulated spectrin-associated protein 1-like 1	CAMSAP1L1	659.7	-1.16934	-1.32
202126_at	PRP4 pre-mRNA processing factor 4 homolog B (yeast)	PRPF4B	282.1	-1.99763	-1.32
213154_s_at	bicaudal D homolog 2 (Drosophila)	BICD2	416.1	-1.59322	-1.32
200983_x_at	CD59 antigen p18-20 (antigen identified by monoclonal antibodies 16.3A5, EJ16, EJ30, EL32 and G344)	CD59	396.1	-1.64217	-1.33
226421_at	Alport syndrome, mental retardation, midface hypoplasia and elliptocytosis chromosomal region, gene 1	AMMECR1	355.4	-1.75247	-1.33
207956_x_at	androgen-induced proliferation inhibitor	APRIN	746.7	-1.06719	-1.33
222620_s_at	DnaJ (Hsp40) homolog, subfamily C, member 1	DNAJC1	458.6	-1.49841	-1.33
202541_at	small inducible cytokine subfamily E, member 1 (endothelial monocyte-activating)	SCYE1	437.1	-1.54487	-1.33
203964_at	N-myc (and STAT) interactor	NMI	388.7	-1.66125	-1.33
204224_s_at	GTP cyclohydrolase 1 (dopa-responsive dystonia)	GCH1	318.9	-1.86588	-1.33
202933_s_at	v-yes-1 Yamaguchi sarcoma viral oncogene homolog 1	YES1	317.0	-1.87212	-1.33
218728_s_at	cornichon homolog 4 (Drosophila)	CNIH4	428.4	-1.56439	-1.33
223051_at	Ssu72 RNA polymerase II CTD phosphatase homolog (yeast)	SSU72	419.9	-1.58424	-1.33
242260_at	Matrin 3	MATR3	428.8	-1.56367	-1.33
37005_at	neuroblastoma, suppression of tumorigenicity 1	NBL1	451.4	-1.51364	-1.33
219010_at	chromosome 1 open reading frame 106	C1orf106	612.8	-1.23264	-1.33
217979_at	Tetraspanin 13	TM4SF13	529.5	-1.36333	-1.33
218139_s_at	chromosome 14 open reading frame 108	C14orf108	536.6	-1.35107	-1.33
218185_s_at	armadillo repeat containing 1	ARMC1	340.2	-1.79782	-1.33
225049_at	biogenesis of lysosome-related organelles complex-1, subunit 2	BLOC1S2	336.9	-1.80813	-1.33
212145_at	mitochondrial ribosomal protein S27	MRPS27	283.3	-1.99307	-1.33
223105_s_at	transmembrane protein 14C /// transmembrane protein 14B	TMEM14C /// TMEM14B	300.3	-1.92981	-1.34
213878_at	RecQ protein-like (DNA helicase Q1-like)	RECQL	324.4	-1.84792	-1.34
218007_s_at	ribosomal protein S27-like	RPS27L	504.4	-1.40826	-1.34
210639_s_at	ATG5 autophagy related 5 homolog (S.	ATG5	481.7	-1.45138	-1.34

	cerevisiae)					
217874_at	succinate-CoA ligase, GDP-forming, alpha subunit	SUCLG1	441.3	-1.53551	-1.34	
229884_s_at	mitochondrial ribosomal protein L2	MRPL2	316.2	-1.87504	-1.34	
223516_s_at	chromosome 6 open reading frame 49	C6orf49	425.3	-1.57157	-1.34	
232681_at	PC4 and SFRS1 interacting protein 1	PSIP1	287.2	-1.97823	-1.34	
205055_at	integrin, alpha E (antigen CD103, human mucosal lymphocyte antigen 1; alpha polypeptide)	ITGAE	511.3	-1.39559	-1.34	
208716_s_at	transmembrane and coiled-coil domains 1	TMCO1	353.9	-1.75704	-1.34	
218911_at	YEATS domain containing 4	YEATS4	529.4	-1.36348	-1.34	
221745_at	WD repeat domain 68	WDR68	335.9	-1.81118	-1.34	
242438_at	---	---	318.9	-1.86593	-1.35	
218167_at	archaemetzincins-2	AMZ2	481.3	-1.45228	-1.35	
226989_at	RGM domain family, member B	RGMB	329.1	-1.83257	-1.35	
219231_at	nuclear receptor coactivator 6 interacting protein	NCOA6IP	332.0	-1.82352	-1.35	
225315_at	mitochondrial ribosomal protein L21	MRPL21	403.8	-1.62301	-1.35	
221434_s_at	chromosome 14 open reading frame 156 /// chromosome 14 open reading frame 156	C14orf156	390.2	-1.65733	-1.35	
37012_at	capping protein (actin filament) muscle Z-line, beta	CAPZB	306.9	-1.90675	-1.35	
203341_at	CCAAT/enhancer binding protein zeta	CEBPZ	474.2	-1.46642	-1.35	
218095_s_at	TPA regulated locus	TPARL	545.8	-1.33557	-1.35	
208447_s_at	phosphoribosyl pyrophosphate synthetase 1	PRPS1	338.6	-1.80280	-1.35	
208764_s_at	ATP synthase, H <sup>+</sup> transporting, mitochondrial F0 complex, subunit c (subunit 9), isoform 2	ATP5G2	475.6	-1.46357	-1.35	
228710_at	---	---	479.3	-1.45625	-1.35	
230793_at	Leucine rich repeat containing 16	LRRC16	307.2	-1.90549	-1.35	
223106_at	transmembrane protein 14C	TMEM14C	337.5	-1.80629	-1.35	
223700_at	GAJ protein	GAJ	412.9	-1.60089	-1.35	
222752_s_at	chromosome 1 open reading frame 75	C1orf75	337.7	-1.80552	-1.35	
219029_at	hypothetical protein FLJ21657	FLJ21657	502.2	-1.41234	-1.35	
226386_at	chromosome 7 open reading frame 30	C7orf30	345.6	-1.78143	-1.35	
212586_at	calpastatin	CAST	446.3	-1.52457	-1.35	
209433_s_at	phosphoribosyl pyrophosphate amidotransferase	PPAT	340.5	-1.79709	-1.35	
209023_s_at	stromal antigen 2	STAG2	489.3	-1.43675	-1.35	
224718_at	YY1 transcription factor	YY1	308.4	-1.90134	-1.36	
208151_x_at	DEAD (Asp-Glu-Ala-Asp) box polypeptide 17 /// DEAD (Asp-Glu-Ala-Asp) box polypeptide 17	DDX17	501.2	-1.41415	-1.36	
213979_s_at	Hypothetical protein LOC285463	CTBP1	295.3	-1.94805	-1.36	
1559038_at	septin 2	2-Sep	407.2	-1.61461	-1.36	

207507_s_at	ATP synthase, H <sup>+</sup> transporting, mitochondrial F0 complex, subunit c (subunit 9) isoform 3	ATP5G3	305.8	-1.91059	-1.36
229235_at	Nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 2 interacting protein	NFATC2IP	288.2	-1.97411	-1.36
218852_at	chromosome 14 open reading frame 10	C14orf10	846.6	-0.96891	-1.36
213015_at	Bobby sox homolog (Drosophila)	BBX	300.2	-1.93027	-1.36
202170_s_at	aminoadipate-semialdehyde dehydrogenase-phosphopantetheinyl transferase	AASDHPPT	352.5	-1.76114	-1.36
225081_s_at	cell division cycle associated 7-like	CDCA7L	355.8	-1.75134	-1.36
201279_s_at	disabled homolog 2, mitogen-responsive phosphoprotein (Drosophila)	DAB2	421.1	-1.58142	-1.36
226190_at	Mitogen-activated protein kinase kinase kinase 13	RPL4	302.0	-1.92390	-1.36
222765_x_at	chromosome 20 open reading frame 6	C20orf6	402.8	-1.62535	-1.36
200799_at	heat shock 70kDa protein 1A	HSPA1A	485.8	-1.44343	-1.36
201036_s_at	L-3-hydroxyacyl-Coenzyme A dehydrogenase, short chain	HADHSC	351.8	-1.76322	-1.36
213376_at	zinc finger and BTB domain containing 1	ZBTB1	435.5	-1.54842	-1.36
222476_at	CCR4-NOT transcription complex, subunit 6	CNOT6	387.6	-1.66394	-1.36
243495_s_at	CDNA FLJ36515 fis, clone TRACH2001810	---	392.4	-1.65170	-1.36
210908_s_at	prefoldin 5	PFDN5	348.9	-1.77170	-1.36
232060_at	Full-length cDNA clone CS0DD009YB17 of Neuroblastoma Cot 50-normalized of Homo sapiens (human)	---	458.9	-1.49763	-1.37
228614_at	hypothetical protein LOC205251	LOC205251	438.1	-1.54260	-1.37
201801_s_at	solute carrier family 29 (nucleoside transporters), member 1	SLC29A1	403.0	-1.62490	-1.37
211165_x_at	EPH receptor B2	EPHB2	370.6	-1.70947	-1.37
1552291_at	phosphatidylinositol glycan, class X	PIGX	547.3	-1.33307	-1.37
224812_at	3-hydroxyisobutyrate dehydrogenase	HIBADH	413.5	-1.59932	-1.37
228332_s_at	chromosome 11 open reading frame 31	C11orf31	450.8	-1.51479	-1.37
202651_at	lysophosphatidylglycerol acyltransferase 1	LPGAT1	338.2	-1.80418	-1.37
228180_at	Smu-1 suppressor of mec-8 and unc-52 homolog (C. elegans)	SMU1	292.1	-1.95984	-1.37
203478_at	NADH dehydrogenase (ubiquinone) 1, subcomplex unknown, 1, 6kDa	NDUFC1	364.2	-1.72731	-1.37
212301_at	Rtf1, Paf1/RNA polymerase II complex component, homolog (S. cerevisiae)	RTF1	340.2	-1.79790	-1.37
226881_at	GrpE-like 2, mitochondrial (E. coli)	GRPEL2	440.5	-1.53735	-1.37
218100_s_at	estrogen-related receptor beta like 1	ESRRBL1	402.8	-1.62546	-1.37
227767_at	casein kinase 1, gamma 3	CSNK1G3	368.7	-1.71479	-1.37
218536_at	MRS2-like, magnesium homeostasis factor	MRS2L	508.2	-1.40122	-1.37

	( <i>S. cerevisiae</i> )					
225658_at	hypothetical protein LOC339745	LOC339745	564.9	-1.30461	-1.37	
218003_s_at	FK506 binding protein 3, 25kDa	FKBP3	329.2	-1.83241	-1.37	
209028_s_at	abl-interactor 1	ABI1	531.5	-1.35977	-1.37	
210788_s_at	dehydrogenase/reductase (SDR family) member 7	DHRS7	301.6	-1.92514	-1.37	
224879_at	chromosome 9 open reading frame 123	C9orf123	480.9	-1.45305	-1.37	
208845_at	voltage-dependent anion channel 3	VDAC3	431.0	-1.55847	-1.37	
229083_at	Heterogeneous nuclear ribonucleoprotein A0	HNRPA0	322.5	-1.85416	-1.37	
222403_at	mitochondrial carrier homolog 2 ( <i>C. elegans</i> )	MTCH2	538.2	-1.34838	-1.37	
227364_at	---	---	394.1	-1.64726	-1.38	
228416_at	activin A receptor, type IIA	ACVR2A	472.0	-1.47080	-1.38	
230291_s_at	Nuclear factor I/B	NFIB	423.4	-1.57592	-1.38	
1554168_a_at	SH3-domain kinase binding protein 1	SH3KBP1	294.8	-1.94995	-1.38	
204387_x_at	mitochondrial ribosomal protein 63	MRP63	332.5	-1.82170	-1.38	
213604_at	Transcription elongation factor B (SIII), polypeptide 3 (110kDa, elongin A)	TCEB3	504.8	-1.40754	-1.38	
204244_s_at	activator of S phase kinase	ASK	327.2	-1.83888	-1.38	
204610_s_at	hepatitis delta antigen-interacting protein A	DIPA	337.3	-1.80674	-1.38	
209162_s_at	PRP4 pre-mRNA processing factor 4 homolog (yeast)	PRPF4	457.7	-1.50018	-1.38	
209817_at	protein phosphatase 3 (formerly 2B), catalytic subunit, beta isoform (calcineurin A beta)	PPP3CB	388.3	-1.66215	-1.38	
212023_s_at	antigen identified by monoclonal antibody Ki-67	MKI67	365.9	-1.72265	-1.38	
226683_at	Sorting nexin associated golgi protein 1	SNAG1	356.2	-1.75028	-1.38	
222627_at	vacuolar protein sorting 54 (yeast)	VPS54	368.1	-1.71645	-1.38	
220232_at	stearoyl-CoA desaturase 5	SCD5	571.2	-1.29464	-1.38	
218149_s_at	zinc finger protein 395	ZNF395	390.1	-1.65759	-1.39	
227556_at	non-metastatic cells 7, protein expressed in (nucleoside-diphosphate kinase)	NME7	305.4	-1.91188	-1.39	
224437_s_at	chromosome 6 open reading frame 55 /// chromosome 6 open reading frame 55	C6orf55	473.7	-1.46738	-1.39	
218556_at	ORM1-like 2 ( <i>S. cerevisiae</i> )	ORMDL2	322.6	-1.85351	-1.39	
208079_s_at	serine/threonine kinase 6	STK6	400.2	-1.63203	-1.39	
36907_at	mevalonate kinase (mevalonic aciduria)	MVK	322.0	-1.85571	-1.39	
201699_at	proteasome (prosome, macropain) 26S subunit, ATPase, 6	PSMC6	515.5	-1.38807	-1.39	
225984_at	protein kinase, AMP-activated, alpha 1 catalytic subunit	PRKAA1	445.1	-1.52722	-1.39	
203494_s_at	translokin	PIG8	412.7	-1.60125	-1.39	
225230_at	transmembrane protein 77	TMEM77	430.8	-1.55906	-1.39	
201486_at	reticulocalbin 2, EF-hand calcium binding	RCN2	434.3	-1.55106	-1.39	

	domain				
46167_at	tetratricopeptide repeat domain 4 /// chromosome 1 open reading frame 175	TTC4 /// C1orf175	466.1	-1.48288	-1.39
209009_at	esterase D/formylglutathione hydrolase	ESD	341.5	-1.79402	-1.39
228726_at	Serpin peptidase inhibitor, clade B (ovalbumin), member 1	SERPINB1	646.9	-1.18606	-1.39
201512_s_at	translocase of outer mitochondrial membrane 70 homolog A (yeast)	TOMM70A	327.5	-1.83764	-1.39
220235_s_at	chromosome 1 open reading frame 103	C1orf103	447.7	-1.52149	-1.39
205527_s_at	gem (nuclear organelle) associated protein 4	GEMIN4	301.7	-1.92478	-1.39
212919_at	DCP2 decapping enzyme homolog (S. cerevisiae)	DCP2	403.8	-1.62306	-1.40
230078_at	KIAA1961 gene	KIAA1961	368.4	-1.71556	-1.40
211962_s_at	zinc finger protein 36, C3H type-like 1	ZFP36L1	404.0	-1.62247	-1.40
224618_at	---	---	549.2	-1.32998	-1.40
201405_s_at	COP9 constitutive photomorphogenic homolog subunit 6 (Arabidopsis)	COPS6	318.2	-1.86807	-1.40
201536_at	dual specificity phosphatase 3 (vaccinia virus phosphatase VH1-related)	DUSP3	321.9	-1.85611	-1.40
220964_s_at	RAB1B, member RAS oncogene family /// RAB1B, member RAS oncogene family	RAB1B	298.2	-1.93744	-1.40
212220_at	proteasome (prosome, macropain) activator subunit 4	PSME4	439.6	-1.53933	-1.40
218862_at	ankyrin repeat and SOCS box-containing 13	ASB13	284.8	-1.98707	-1.40
222811_at	hypothetical protein FLJ11171	FLJ11171	769.8	-1.04282	-1.40
225036_at	Src homology 2 domain containing adaptor protein B /// chromosome 9 open reading frame 105	SHB /// C9orf105	343.0	-1.78936	-1.40
228760_at	Splicing factor, arginine/serine-rich, 46kD	SRP46	554.9	-1.32072	-1.40
204373_s_at	centrosome-associated protein 350	CAP350	529.5	-1.36339	-1.40
1553587_a_at	polymerase (DNA-directed), epsilon 4 (p12 subunit)	POLE4	321.1	-1.85863	-1.40
202077_at	NADH dehydrogenase (ubiquinone) 1, alpha/beta subcomplex, 1, 8kDa	NDUFAB1	460.6	-1.49406	-1.40
207014_at	gamma-aminobutyric acid (GABA) A receptor, alpha 2	GABRA2	494.8	-1.42615	-1.40
218496_at	ribonuclease H1	RNASEH1	390.4	-1.65675	-1.41
209366_x_at	cytochrome b-5	CYB5	296.2	-1.94458	-1.41
202594_at	leptin receptor overlapping transcript-like 1	LEPROTL1	319.2	-1.86486	-1.41
227126_at	Transcribed locus	---	662.2	-1.16616	-1.41
210681_s_at	ubiquitin specific peptidase 15	USP15	514.3	-1.39021	-1.41
205134_s_at	nuclear fragile X mental retardation protein interacting protein 1	NUFIP1	622.4	-1.21915	-1.41
218930_s_at	hypothetical protein FLJ11273	FLJ11273	342.5	-1.79102	-1.41

213795_s_at	protein tyrosine phosphatase, receptor type, A	PTPRA	376.5	-1.69344	-1.41
203023_at	hypothetical protein HSPC111	HSPC111	479.8	-1.45526	-1.41
204201_s_at	protein tyrosine phosphatase, non-receptor type 13 (APO-1/CD95 (Fas)-associated phosphatase)	PTPN13	456.7	-1.50236	-1.41
229075_at	Transcribed locus	---	349.2	-1.77081	-1.41
225974_at	transmembrane protein 64	TMEM64	484.4	-1.44625	-1.41
213698_at	zinc finger, MYM-type 6	ZMYM6	334.7	-1.81482	-1.41
206055_s_at	small nuclear ribonucleoprotein polypeptide A'	SNRPA1	339.3	-1.80078	-1.41
204318_s_at	G-2 and S-phase expressed 1	GTSE1	408.9	-1.61038	-1.41
202587_s_at	adenylate kinase 1	AK1	343.0	-1.78943	-1.41
205690_s_at	G10 protein	G10	286.0	-1.98240	-1.41
228249_at	HEPIS	LOC119710	595.4	-1.25782	-1.42
1569110_x_at	programmed cell death 6	PDCD6	389.4	-1.65952	-1.42
204291_at	zinc finger protein 518	ZNF518	358.6	-1.74329	-1.42
204955_at	sushi-repeat-containing protein, X-linked	SRPX	387.3	-1.66492	-1.42
244261_at	interleukin 28 receptor, alpha (interferon, lambda receptor)	IL28RA	418.6	-1.58715	-1.42
224865_at	male sterility domain containing 2	MLSTD2	407.1	-1.61497	-1.42
222391_at	transmembrane protein 30A	TMEM30A	490.7	-1.43394	-1.42
200975_at	palmitoyl-protein thioesterase 1 (ceroid-lipofuscinosis, neuronal 1, infantile)	PPT1	424.3	-1.57404	-1.42
226505_x_at	ubiquitin specific peptidase 32	USP32	302.7	-1.92147	-1.42
208835_s_at	cisplatin resistance-associated overexpressed protein	CROP	286.8	-1.97946	-1.42
209233_at	C2f protein	C2F	478.4	-1.45801	-1.42
213483_at	peptidylprolyl isomerase domain and WD repeat containing 1	PPWD1	322.7	-1.85350	-1.42
205251_at	period homolog 2 (Drosophila)	PER2	438.6	-1.54141	-1.42
219283_at	C1GALT1-specific chaperone 1	C1GALT1C1	681.7	-1.14178	-1.42
226245_at	potassium channel tetramerisation domain containing 1	KCTD1	392.4	-1.65165	-1.42
206323_x_at	oligophrenin 1	OPHN1	532.8	-1.35769	-1.42
208882_s_at	E3 ubiquitin protein ligase, HECT domain containing, 1	EDD1	587.8	-1.26917	-1.43
232914_s_at	synaptotagmin-like 2	SYTL2	396.9	-1.64028	-1.43
201980_s_at	Ras suppressor protein 1	RSU1	364.4	-1.72695	-1.43
214600_at	TEA domain family member 1 (SV40 transcriptional enhancer factor)	TEAD1	318.9	-1.86600	-1.43
224445_s_at	zinc finger, FYVE domain containing 21 /// zinc finger, FYVE domain containing 21	ZFYVE21	398.8	-1.63548	-1.43
221753_at	slingshot homolog 1 (Drosophila)	SSH1	424.1	-1.57435	-1.43
220526_s_at	mitochondrial ribosomal protein L20	MRPL20	485.8	-1.44341	-1.43

210986_s_at	tropomyosin 1 (alpha)	TPM1	298.4	-1.93688	-1.43
228680_at	Full-length cDNA clone CS0CAP007YI15 of Thymus of Homo sapiens (human)	---	548.8	-1.33064	-1.43
229630_s_at	Wilms tumor 1 associated protein	WTAP	523.5	-1.37376	-1.43
209274_s_at	HESB like domain containing 2	HBLD2	553.8	-1.32252	-1.43
212615_at	chromodomain helicase DNA binding protein 9	CHD9	429.9	-1.56108	-1.43
208541_x_at	Transcription factor A, mitochondrial	TFAM	505.3	-1.40650	-1.43
218233_s_at	chromosome 6 open reading frame 49	C6orf49	366.6	-1.72079	-1.43
202916_s_at	family with sequence similarity 20, member B	FAM20B	513.2	-1.39213	-1.43
222477_s_at	transmembrane 7 superfamily member 3	TM7SF3	416.8	-1.59158	-1.43
44790_s_at	chromosome 13 open reading frame 18	C13orf18	544.5	-1.33772	-1.44
207431_s_at	degenerative spermatocyte homolog 1, lipid desaturase (Drosophila)	DEGS1	561.6	-1.30992	-1.44
203259_s_at	HD domain containing 2	HDDC2	511.1	-1.39599	-1.44
224596_at	solute carrier family 44, member 1	SLC44A1	452.4	-1.51140	-1.44
213980_s_at	C-terminal binding protein 1	CTBP1	291.1	-1.96342	-1.44
230466_s_at	Mesenchymal stem cell protein DSC96	---	471.2	-1.47244	-1.44
218147_s_at	glycosyltransferase 8 domain containing 1	GLT8D1	468.7	-1.47749	-1.44
212688_at	phosphoinositide-3-kinase, catalytic, beta polypeptide	PIK3CB	453.2	-1.50980	-1.44
1553015_a_at	RecQ protein-like 4	RECQL4	417.8	-1.58924	-1.44
218830_at	ribosomal protein L26-like 1	RPL26L1	412.9	-1.60079	-1.44
222621_at	DnaJ (Hsp40) homolog, subfamily C, member 1	DNAJC1	608.5	-1.23878	-1.45
200667_at	ubiquitin-conjugating enzyme E2D 3 (UBC4/5 homolog, yeast)	UBE2D3	556.2	-1.31856	-1.45
226374_at	Full-length cDNA clone CS0DF012YG01 of Fetal brain of Homo sapiens (human)	---	619.9	-1.22271	-1.45
225297_at	coiled-coil domain containing 5 (spindle associated)	CCDC5	516.0	-1.38704	-1.45
226895_at	Nuclear factor I/C (CCAAT-binding transcription factor)	NFIC	419.1	-1.58597	-1.45
221196_x_at	chromosome X open reading frame 53	CXorf53	523.1	-1.37451	-1.45
209173_at	anterior gradient 2 homolog (Xenopus laevis)	AGR2	574.2	-1.29001	-1.45
217898_at	chromosome 15 open reading frame 24	C15orf24	382.0	-1.67884	-1.45
202911_at	mutS homolog 6 (E. coli)	MSH6	298.5	-1.93646	-1.45
205928_at	zinc finger protein 443	ZNF443	383.6	-1.67446	-1.45
218671_s_at	ATPase inhibitory factor 1	ATPIF1	425.8	-1.57041	-1.45
226132_s_at	mannosidase, endo-alpha-like	MANEAL	350.2	-1.76788	-1.45
224932_at	chromosome 22 open reading frame 16	C22orf16	291.7	-1.96125	-1.45
201876_at	paraoxonase 2	PON2	535.5	-1.35295	-1.46
202797_at	SAC1 suppressor of actin mutations 1-like (yeast)	SACM1L	510.9	-1.39623	-1.46

218160_at	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 8, 19kDa	NDUFA8	686.3	-1.13616	-1.46
225553_at	CDNA FLJ12874 fis, clone NT2RP2003769	---	372.7	-1.70382	-1.46
209662_at	centrin, EF-hand protein, 3 (CDC31 homolog, yeast)	CETN3	366.4	-1.72131	-1.46
221705_s_at	hypothetical protein FLJ21168 /// hypothetical protein FLJ21168	FLJ21168	464.9	-1.48524	-1.46
222603_at	KIAA1815	KIAA1815	528.3	-1.36543	-1.46
202605_at	glucuronidase, beta	GUSB	335.4	-1.81273	-1.46
202961_s_at	ATP synthase, H <sup>+</sup> transporting, mitochondrial F0 complex, subunit f, isoform 2	ATP5J2	540.3	-1.34489	-1.46
209303_at	NADH dehydrogenase (ubiquinone) Fe-S protein 4, 18kDa (NADH-coenzyme Q reductase)	NDUFS4	297.6	-1.93948	-1.46
204808_s_at	transmembrane protein 5	TMEM5	467.5	-1.47989	-1.46
205771_s_at	A kinase (PRKA) anchor protein 7	AKAP7	559.5	-1.31324	-1.46
209340_at	UDP-N-acteylglucosamine pyrophosphorylase 1	UAP1	288.0	-1.97506	-1.46
224830_at	nudix (nucleoside diphosphate linked moiety X)-type motif 21	NUDT21	292.0	-1.95997	-1.46
220945_x_at	MANSC domain containing 1	MANSC1	634.8	-1.20217	-1.46
229265_at	Atrophin 1	DRPLA	555.7	-1.31944	-1.46
225893_at	Clone TESTIS-724 mRNA sequence	---	319.6	-1.86348	-1.46
203048_s_at	KIAA0372	KIAA0372	610.4	-1.23612	-1.47
228745_at	Hypothetical protein FLJ13611	FLJ13611	289.3	-1.97023	-1.47
224345_x_at	growth and transformation-dependent protein /// growth and transformation-dependent protein	E2IG5	411.5	-1.60417	-1.47
219549_s_at	reticulon 3	RTN3	411.2	-1.60497	-1.47
211730_s_at	polymerase (RNA) II (DNA directed) polypeptide L, 7.6kDa /// polymerase (RNA) II (DNA directed) polypeptide L, 7.6kDa	POLR2L	422.6	-1.57797	-1.47
203022_at	ribonuclease H2, large subunit	RNASEH2A	408.2	-1.61224	-1.47
1559822_s_at	Homo sapiens, clone IMAGE:3606519	---	303.5	-1.91863	-1.47
224332_s_at	mitochondrial ribosomal protein L43 /// mitochondrial ribosomal protein L43	MRPL43	497.6	-1.42094	-1.47
236327_at	Choline/ethanolamine phosphotransferase 1	CEPT1	473.1	-1.46860	-1.47
212714_at	La ribonucleoprotein domain family, member 4	LARP4	503.7	-1.40953	-1.47
1555446_s_at	transmembrane protein 1	TMEM1	539.5	-1.34624	-1.47
205191_at	retinitis pigmentosa 2 (X-linked recessive)	RP2	764.3	-1.04848	-1.47
212297_at	ATPase type 13A3	ATP13A3	413.0	-1.60047	-1.47
203209_at	replication factor C (activator 1) 5, 36.5kDa	RFC5	489.1	-1.43702	-1.47

201153_s_at	muscleblind-like (Drosophila)	MBNL1	376.2	-1.69427	-1.47
225777_at	chromosome 9 open reading frame 140	C9orf140	461.4	-1.49241	-1.47
203588_s_at	transcription factor Dp-2 (E2F dimerization partner 2)	TFDP2	416.5	-1.59226	-1.47
201757_at	NADH dehydrogenase (ubiquinone) Fe-S protein 5, 15kDa (NADH-coenzyme Q reductase)	NDUFS5	588.4	-1.26833	-1.47
226885_at	Transcribed locus	---	532.7	-1.35785	-1.47
225143_at	sideroflexin 4	SFXN4	441.6	-1.53488	-1.47
226924_at	hypothetical gene supported by BC036588	LOC400657	358.0	-1.74493	-1.47
219296_at	zinc finger, DHHC-type containing 13	ZDHHC13	295.1	-1.94880	-1.47
225760_at	myb-like, SWIRM and MPN domains 1	MYSM1	602.4	-1.24763	-1.47
1553956_at	amyotrophic lateral sclerosis 2 (juvenile) chromosome region, candidate 4	ALS2CR4	499.7	-1.41690	-1.48
204185_x_at	peptidylprolyl isomerase D (cyclophilin D)	PPID	324.3	-1.84824	-1.48
209646_x_at	aldehyde dehydrogenase 1 family, member B1	ALDH1B1	512.6	-1.39318	-1.48
213606_s_at	Rho GDP dissociation inhibitor (GDI) alpha	ARHGDI1	393.3	-1.64929	-1.48
235033_at	Aminopeptidase-like 1	NPEPL1	480.1	-1.45468	-1.48
234000_s_at	butyrate-induced transcript 1	HSPC121	656.2	-1.17387	-1.48
201462_at	secernin 1	SCRN1	391.8	-1.65331	-1.48
200761_s_at	ADP-ribosylation-like factor 6 interacting protein 5	ARL6IP5	484.9	-1.44524	-1.48
212538_at	dedicator of cytokinesis 9	DOCK9	473.0	-1.46873	-1.48
202950_at	crystallin, zeta (quinone reductase)	CRYZ	409.4	-1.60922	-1.48
218499_at	Mst3 and SOK1-related kinase	MASK	523.7	-1.37342	-1.48
224824_at	family with sequence similarity 36, member A	FAM36A	387.1	-1.66543	-1.48
201026_at	eukaryotic translation initiation factor 5B	EIF5B	337.4	-1.80654	-1.48
213879_at	SMT3 suppressor of mif two 3 homolog 2 (yeast)	SUMO2	434.6	-1.55034	-1.48
238465_at	hypothetical protein MGC33648	MGC33648	780.4	-1.03196	-1.48
229215_at	achaete-scute complex-like 2 (Drosophila)	ASCL2	463.7	-1.48765	-1.48
229174_at	---	---	633.0	-1.20468	-1.48
224308_s_at	KIAA1287	KIAA1287	340.0	-1.79864	-1.48
201930_at	MCM6 minichromosome maintenance deficient 6 (MIS5 homolog, S. pombe) (S. cerevisiae)	MCM6	409.4	-1.60928	-1.48
202850_at	ATP-binding cassette, sub-family D (ALD), member 3	ABCD3	309.0	-1.89946	-1.48
225649_s_at	serine/threonine kinase 35	STK35	527.6	-1.36667	-1.49
212677_s_at	KIAA0582	KIAA0582	438.9	-1.54086	-1.49
222077_s_at	Rac GTPase activating protein 1	RACGAP1	352.6	-1.76061	-1.49
233898_s_at	FGFR1 oncogene partner 2	FGFR1OP2	444.9	-1.52764	-1.49

211752_s_at	NADH dehydrogenase (ubiquinone) Fe-S protein 7, 20kDa (NADH-coenzyme Q reductase) /// NADH dehydrogenase (ubiquinone) Fe-S protein 7, 20kDa (NADH-coenzyme Q reductase)	NDUFS7	339.7	-1.79934	-1.49
221954_at	Chromosome 20 open reading frame 111	C20orf111	386.2	-1.66777	-1.49
203406_at	microfibrillar-associated protein 1	MFAP1	298.6	-1.93597	-1.49
218088_s_at	Ras-related GTP binding C	RRAGC	510.1	-1.39774	-1.49
223716_s_at	zinc finger protein 265	ZNF265	344.0	-1.78636	-1.49
201817_at	ubiquitin protein ligase E3C	UBE3C	460.6	-1.49410	-1.49
219137_s_at	chromosome 2 open reading frame 33	C2orf33	363.3	-1.73008	-1.49
211698_at	CREBBP/EP300 inhibitor 1 /// CREBBP/EP300 inhibitor 1	CRI1	318.8	-1.86626	-1.49
227884_at	TAF15 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 68kDa	TAF15	473.4	-1.46789	-1.49
201824_at	ring finger protein 14	RNF14	455.3	-1.50534	-1.49
215867_x_at	carbonic anhydrase XII	CA12	328.2	-1.83540	-1.49
218034_at	tetratricopeptide repeat domain 11	TTC11	287.6	-1.97644	-1.49
203285_s_at	heparan sulfate 2-O-sulfotransferase 1	HS2ST1	550.8	-1.32729	-1.49
221229_s_at	hypothetical protein FLJ20628 /// hypothetical protein FLJ20628	FLJ20628	450.6	-1.51525	-1.49
202213_s_at	cullin 4B	CUL4B	392.0	-1.65282	-1.49
203714_s_at	tubulin-specific chaperone e	TBCE	487.3	-1.44059	-1.50
235072_s_at	Transcribed locus	---	631.1	-1.20719	-1.50
218187_s_at	chromosome 8 open reading frame 33	C8orf33	452.2	-1.51196	-1.50
223271_s_at	CTD (carboxy-terminal domain, RNA polymerase II, polypeptide A) small phosphatase like 2	CTDSPL2	425.7	-1.57079	-1.50
218756_s_at	short-chain dehydrogenase/reductase	MGC4172	444.0	-1.52968	-1.50
221540_x_at	general transcription factor IIIH, polypeptide 2, 44kDa	GTF2H2	434.2	-1.55139	-1.50
202171_at	zinc finger protein 161	ZNF161	459.0	-1.49747	-1.50
226816_s_at	KIAA1143	KIAA1143	558.3	-1.31509	-1.50
229360_at	suppressor of hairy wing homolog 2 (Drosophila)	SUHW2	477.8	-1.45917	-1.50
212058_at	U2-associated SR140 protein	SR140	439.2	-1.54025	-1.50
201029_s_at	CD99 antigen	CD99	527.7	-1.36637	-1.50
221943_x_at	Ribosomal protein L38	RPL38	467.3	-1.48042	-1.50
202118_s_at	copine III	CPNE3	337.4	-1.80649	-1.50
223977_s_at	chromosome 18 open reading frame 2	C18orf2	283.2	-1.99310	-1.50
225711_at	ADP-ribosylation-like factor 6 interacting protein 6	ARL6IP6	721.1	-1.09539	-1.50
201675_at	A kinase (PRKA) anchor protein 1	AKAP1	450.5	-1.51550	-1.50
212959_s_at	N-acetylglucosamine-1-phosphate transferase, alpha and beta subunits	GNPTAB	311.0	-1.89234	-1.50

209029_at	COP9 constitutive photomorphogenic homolog subunit 7A (Arabidopsis)	COPS7A	613.9	-1.23106	-1.50
205047_s_at	asparagine synthetase	ASNS	380.8	-1.68198	-1.51
226337_at	SCY1-like 1 binding protein 1	SCYL1BP1	362.1	-1.73336	-1.51
204559_s_at	LSM7 homolog, U6 small nuclear RNA associated (S. cerevisiae)	LSM7	407.0	-1.61520	-1.51
213350_at	Ribosomal protein S11	RPS11	340.3	-1.79762	-1.51
213517_at	Poly(rC) binding protein 2	PCBP2	532.0	-1.35893	-1.51
226181_at	tubulin, epsilon 1	TUBE1	388.8	-1.66091	-1.51
201724_s_at	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 1 (GalNAc-T1)	GALNT1	481.2	-1.45252	-1.51
224766_at	ribosomal protein L37	RPL37	459.2	-1.49712	-1.51
231192_at	Endothelial differentiation, lysophosphatidic acid G-protein-coupled receptor, 7	EDG7	732.4	-1.08277	-1.51
238122_at	RNA binding motif protein 12B	RBM12B	503.2	-1.41037	-1.51
225116_at	Homeodomain interacting protein kinase 2	HIPK2	608.9	-1.23821	-1.51
201687_s_at	apoptosis inhibitor 5	API5	317.6	-1.87020	-1.51
225614_at	hypothetical protein BC012010	LOC113174	338.7	-1.80247	-1.51
201242_s_at	ATPase, Na <sup>+</sup> /K <sup>+</sup> transporting, beta 1 polypeptide	ATP1B1	307.3	-1.90535	-1.51
212741_at	monoamine oxidase A	MAOA	447.4	-1.52217	-1.51
1554351_a_at	TIP41, TOR signalling pathway regulator-like (S. cerevisiae)	TIPRL	523.4	-1.37388	-1.51
235144_at	RAS and EF-hand domain containing	RASEF	708.8	-1.10946	-1.51
209666_s_at	conserved helix-loop-helix ubiquitous kinase	CHUK	326.7	-1.84044	-1.51
205609_at	angiopoietin 1	ANGPT1	799.1	-1.01342	-1.51
225501_at	PHD finger protein 6	PHF6	437.5	-1.54402	-1.51
219347_at	nudix (nucleoside diphosphate linked moiety X)-type motif 15	NUDT15	392.7	-1.65095	-1.51
230244_at	ASCL830	UNQ830	290.5	-1.96576	-1.51
223005_s_at	chromosome 9 open reading frame 5	C9orf5	302.0	-1.92401	-1.51
230669_at	RAS p21 protein activator 2	RASA2	501.9	-1.41280	-1.51
214305_s_at	splicing factor 3b, subunit 1, 155kDa	SF3B1	472.1	-1.47049	-1.51
226600_at	SMILE protein	SMILE	541.9	-1.34208	-1.52
201291_s_at	topoisomerase (DNA) II alpha 170kDa	TOP2A	459.7	-1.49601	-1.52
203291_at	CCR4-NOT transcription complex, subunit 4	CNOT4	566.0	-1.30278	-1.52
225098_at	Abl interactor 2	ABI2	393.9	-1.64784	-1.52
225535_s_at	translocase of inner mitochondrial membrane 23 homolog (yeast)	TIMM23	495.6	-1.42468	-1.52
206364_at	kinesin family member 14	KIF14	406.5	-1.61638	-1.52
226896_at	coiled-coil-helix-coiled-coil-helix domain	CHCHD1	474.9	-1.46492	-1.52

	containing 1				
229092_at	Full length insert cDNA clone YX37E06	---	496.5	-1.42296	-1.52
213102_at	ARP3 actin-related protein 3 homolog (yeast)	ACTR3	441.2	-1.53565	-1.52
201581_at	thioredoxin domain containing 13	TXNDC13	541.0	-1.34369	-1.52
225330_at	hypothetical protein MGC18216	MGC18216	617.5	-1.22596	-1.52
224875_at	hypothetical protein FLJ37562	FLJ37562	517.4	-1.38461	-1.52
213720_s_at	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4	SMARCA4	487.4	-1.44031	-1.53
1555851_s_at	selenoprotein W, 1	SEPW1	311.3	-1.89135	-1.53
209645_s_at	aldehyde dehydrogenase 1 family, member B1	ALDH1B1	439.9	-1.53864	-1.53
224810_s_at	ankyrin repeat domain 13	ANKRD13	385.9	-1.66842	-1.53
222789_at	round spermatid basic protein 1	RSBN1	315.7	-1.87646	-1.53
230413_s_at	Adaptor-related protein complex 1, sigma 2 subunit	AP1S2	591.2	-1.26414	-1.53
238761_at	Mediator of RNA polymerase II transcription, subunit 28 homolog (yeast)	MED28	724.9	-1.09108	-1.53
214864_s_at	glyoxylate reductase/hydroxypyruvate reductase	GRHPR	706.1	-1.11259	-1.53
226713_at	chromosome 3 open reading frame 6	C3orf6	743.0	-1.07113	-1.53
218085_at	chromatin modifying protein 5	CHMP5	590.4	-1.26533	-1.53
218616_at	PHD finger protein 22	PHF22	629.4	-1.20945	-1.53
226742_at	Transcribed locus, moderately similar to XP_512541.1 PREDICTED: similar to hypothetical protein [Pan troglodytes]	---	821.3	-0.99216	-1.53
213504_at	COP9 constitutive photomorphogenic homolog subunit 6 (Arabidopsis)	COPS6	708.0	-1.11046	-1.53
208920_at	sorcin	SRI	483.0	-1.44891	-1.53
222734_at	tryptophanyl tRNA synthetase 2 (mitochondrial)	WARS2	513.5	-1.39167	-1.53
229804_x_at	COBW domain containing 1 /// COBW domain containing 2 /// COBW domain containing 3	CBWD1 /// CBWD2 /// CBWD3	486.8	-1.44150	-1.53
217985_s_at	bromodomain adjacent to zinc finger domain, 1A	BAZ1A	636.6	-1.19968	-1.53
201597_at	cytochrome c oxidase subunit VIIa polypeptide 2 (liver)	COX7A2	392.5	-1.65136	-1.53
229644_at	Prolyl endopeptidase	PREP	652.1	-1.17913	-1.53
208091_s_at	EGFR-coamplified and overexpressed protein /// EGFR-coamplified and overexpressed protein	ECOP	406.4	-1.61664	-1.53
203735_x_at	PTPRF interacting protein, binding protein 1 (liprin beta 1)	PPFIBP1	660.7	-1.16811	-1.53
209337_at	PC4 and SFRS1 interacting protein 1	PSIP1	309.3	-1.89832	-1.54
202581_at	heat shock 70kDa protein 1B	HSPA1B	712.2	-1.10559	-1.54

235346_at	FUN14 domain containing 1	FUNDC1	632.7	-1.20500	-1.54
225007_at	Ras-GTPase-activating protein SH3-domain-binding protein	G3BP	367.8	-1.71734	-1.54
203177_x_at	transcription factor A, mitochondrial	TFAM	509.4	-1.39908	-1.54
219598_s_at	RWD domain containing 1	RWDD1	341.0	-1.79537	-1.54
222404_x_at	butyrate-induced transcript 1	HSPC121	613.3	-1.23194	-1.54
212637_s_at	WW domain containing E3 ubiquitin protein ligase 1	WWP1	716.4	-1.10076	-1.54
202427_s_at	brain protein 44	BRP44	628.7	-1.21045	-1.54
228293_at	novel 58.3 KDA protein	LOC91614	363.9	-1.72829	-1.54
209272_at	NGFI-A binding protein 1 (EGR1 binding protein 1)	NAB1	523.6	-1.37360	-1.54
200849_s_at	S-adenosylhomocysteine hydrolase-like 1	AHCYL1	367.7	-1.71750	-1.54
200655_s_at	calmodulin 1 (phosphorylase kinase, delta)	CALM1	381.2	-1.68097	-1.54
223075_s_at	chromosome 9 open reading frame 58	C9orf58	486.0	-1.44304	-1.54
209903_s_at	ataxia telangiectasia and Rad3 related	ATR	448.1	-1.52058	-1.55
219060_at	chromosome 8 open reading frame 32	C8orf32	293.0	-1.95639	-1.55
228841_at	hypothetical protein LOC90624	LOC90624	533.8	-1.35595	-1.55
218583_s_at	DCN1, defective in cullin neddylation 1, domain containing 1 (S. cerevisiae)	DCUN1D1	357.5	-1.74649	-1.55
221510_s_at	glutaminase	GLS	540.4	-1.34472	-1.55
213128_s_at	ubiquitin protein ligase E3A (human papilloma virus E6-associated protein, Angelman syndrome)	UBE3A	563.4	-1.30698	-1.55
212453_at	KIAA1279	KIAA1279	340.8	-1.79619	-1.55
218592_s_at	cat eye syndrome chromosome region, candidate 5	CECR5	369.9	-1.71147	-1.55
226894_at	solute carrier family 35 (UDP-N-acetylglucosamine (UDP-GlcNAc) transporter), member A3	SLC35A3	453.9	-1.50821	-1.55
204404_at	solute carrier family 12 (sodium/potassium/chloride transporters), member 2	SLC12A2	557.2	-1.31696	-1.55
227850_x_at	CDC42 effector protein (Rho GTPase binding) 5	CDC42EP5	785.7	-1.02665	-1.55
242920_at	KIAA0999 protein	KIAA0999	559.9	-1.31264	-1.55
227286_at	hypothetical protein FLJ90652	FLJ90652	716.2	-1.10092	-1.55
203300_x_at	adaptor-related protein complex 1, sigma 2 subunit	AP1S2	797.1	-1.01530	-1.55
226406_at	chromosome 18 open reading frame 25	C18orf25	793.5	-1.01893	-1.56
225497_at	arginyltransferase 1	ATE1	335.1	-1.81365	-1.56
203007_x_at	lysophospholipase I	LYPLA1	701.5	-1.11801	-1.56
216973_s_at	homeo box B7	HOXB7	436.7	-1.54583	-1.56
203188_at	UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 6	B3GNT6	295.6	-1.94680	-1.56
202300_at	hepatitis B virus x interacting protein	HBXIP	549.0	-1.33028	-1.56

219507_at	arginine/serine-rich coiled-coil 1	RSRC1	552.7	-1.32429	-1.56
214093_s_at	far upstream element (FUSE) binding protein 1	FUBP1	674.9	-1.15018	-1.56
222465_at	chromosome 15 open reading frame 15 /// similar to Ribosomal protein L24-like	C15orf15 /// LOC284288	379.2	-1.68628	-1.56
219295_s_at	procollagen C-endopeptidase enhancer 2	PCOLCE2	596.6	-1.25607	-1.56
221749_at	YTH domain family, member 3	YTHDF3	707.9	-1.11058	-1.56
218411_s_at	MAP3K12 binding inhibitory protein 1	MBIP	719.3	-1.09747	-1.56
218528_s_at	ring finger protein 38	RNF38	826.8	-0.98702	-1.56
239231_at	CDNA FLJ41910 fis, clone PEBLM2007834	---	350.1	-1.76808	-1.57
1558515_at	CDNA FLJ33139 fis, clone UTERU1000109	---	370.2	-1.71067	-1.57
1567014_s_at	nuclear factor (erythroid-derived 2)-like 2	NFE2L2	414.8	-1.59621	-1.57
32723_at	cleavage stimulation factor, 3' pre-RNA, subunit 1, 50kDa	CSTF1	580.3	-1.28059	-1.57
203627_at	Insulin-like growth factor 1 receptor	IGF1R	588.2	-1.26858	-1.57
201941_at	carboxypeptidase D	CPD	504.5	-1.40798	-1.57
212927_at	SMC5 structural maintenance of chromosomes 5-like 1 (yeast)	SMC5L1	415.6	-1.59448	-1.57
226954_at	ubiquitin-conjugating enzyme E2R 2	UBE2R2	496.1	-1.42373	-1.57
214097_at	ribosomal protein S21	RPS21	366.4	-1.72131	-1.57
223016_x_at	zinc finger protein 265	ZNF265	756.3	-1.05685	-1.57
40465_at	DEAD (Asp-Glu-Ala-Asp) box polypeptide 23	DDX23	476.2	-1.46243	-1.57
201237_at	capping protein (actin filament) muscle Z-line, alpha 2	CAPZA2	804.8	-1.00782	-1.57
209628_at	nuclear transport factor 2-like export factor 2	NXT2	764.4	-1.04841	-1.57
212416_at	secretory carrier membrane protein 1	SCAMP1	422.5	-1.57810	-1.57
211478_s_at	dipeptidylpeptidase 4 (CD26, adenosine deaminase complexing protein 2)	DPP4	332.4	-1.82202	-1.57
237992_at	Transcribed locus	---	646.1	-1.18706	-1.58
217810_x_at	leucyl-tRNA synthetase	LARS	440.6	-1.53704	-1.58
209307_at	SWAP-70 protein	SWAP70	564.9	-1.30464	-1.58
210296_s_at	peroxisomal membrane protein 3, 35kDa (Zellweger syndrome)	PXMP3	914.7	-0.91109	-1.58
224580_at	Solute carrier family 38, member 1	SLC38A1	459.2	-1.49708	-1.58
224800_at	WD repeat and FYVE domain containing 1	WDFY1	375.0	-1.69768	-1.58
226472_at	peptidylprolyl isomerase (cyclophilin)-like 4	PPIL4	492.6	-1.43033	-1.58
205202_at	protein-L-isoaspartate (D-aspartate) O-methyltransferase	PCMT1	784.4	-1.02797	-1.58
201667_at	gap junction protein, alpha 1, 43kDa (connexin 43)	GJA1	634.8	-1.20213	-1.58
225667_s_at	family with sequence similarity 84, member A /// hypothetical LOC400944	FAM84A /// LOC400944	479.0	-1.45676	-1.58

241734_at	serum response factor binding protein 1	SRFBP1	575.0	-1.28874	-1.58
200046_at	defender against cell death 1 /// defender against cell death 1	DAD1	530.1	-1.36233	-1.59
229751_s_at	hypothetical protein DKFZp434G1415	DKFZP434G1415	707.7	-1.11072	-1.59
230734_x_at	Striatin, calmodulin binding protein	STRN	331.6	-1.82463	-1.59
223465_at	collagen, type IV, alpha 3 (Goodpasture antigen) binding protein	COL4A3BP	652.4	-1.17878	-1.59
221606_s_at	nucleosomal binding protein 1	NSBP1	634.7	-1.20237	-1.59
228189_at	BCL2-associated athanogene 4	BAG4	614.3	-1.23061	-1.59
214109_at	LPS-responsive vesicle trafficking, beach and anchor containing	LRBA	972.7	-0.86662	-1.59
225194_at	pleiotropic regulator 1 (PRL1 homolog, Arabidopsis)	PLRG1	430.4	-1.56003	-1.59
210532_s_at	chromosome 14 open reading frame 2	C14orf2	529.1	-1.36401	-1.59
222414_at	myeloid/lymphoid or mixed-lineage leukemia 3	MLL3	371.9	-1.70596	-1.59
228543_at	CSRP2 binding protein	CSRP2BP	371.9	-1.70597	-1.59
235113_at	peptidylprolyl isomerase (cyclophilin)-like 5	PPIL5	579.0	-1.28251	-1.59
225170_at	WD repeat domain 5	WDR5	384.9	-1.67106	-1.59
201555_at	MCM3 minichromosome maintenance deficient 3 (S. cerevisiae)	MCM3	399.8	-1.63284	-1.59
218026_at	HSPC009 protein	HSPC009	389.2	-1.66003	-1.59
213599_at	Opa interacting protein 5	OIP5	625.9	-1.21428	-1.60
223334_at	hypothetical protein DKFZp586C1924	DKFZp586C1924	736.5	-1.07824	-1.60
226432_at	MRNA; cDNA DKFZp566C034 (from clone DKFZp566C034)	---	381.0	-1.68152	-1.60
203371_s_at	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 3, 12kDa	NDUFB3	560.4	-1.31183	-1.60
222555_s_at	mitochondrial ribosomal protein L44	MRPL44	563.3	-1.30711	-1.60
226366_at	SNF2 histone linker PHD RING helicase	SHPRH	896.5	-0.92590	-1.60
231370_at	Protein phosphatase 1A (formerly 2C), magnesium-dependent, alpha isoform	PPM1A	528.3	-1.36534	-1.60
221770_at	ribulose-5-phosphate-3-epimerase	RPE	333.2	-1.81976	-1.60
201018_at	eukaryotic translation initiation factor 1A, X-linked	EIF1AX	357.0	-1.74784	-1.60
221492_s_at	ATG3 autophagy related 3 homolog (S. cerevisiae)	ATG3	355.9	-1.75114	-1.60
212781_at	retinoblastoma binding protein 6	RBBP6	684.9	-1.13790	-1.60
224160_s_at	acyl-Coenzyme A dehydrogenase family, member 9	ACAD9	352.2	-1.76186	-1.60
202749_at	tryptophan rich basic protein	WRB	700.7	-1.11894	-1.60
204831_at	Cyclin-dependent kinase 8	CDK8	676.1	-1.14864	-1.60
222000_at	hypothetical protein LOC339448	LOC339448	288.2	-1.97425	-1.60
228069_at	family with sequence similarity 54,	FAM54A	542.3	-1.34148	-1.60

	member A				
224867_at	chromosome 1 open reading frame 151	C1orf151	499.4	-1.41753	-1.60
217099_s_at	gem (nuclear organelle) associated protein 4	GEMIN4	427.9	-1.56562	-1.60
224302_s_at	mitochondrial ribosomal protein S36	MRPS36	339.3	-1.80073	-1.61
242323_at	---	---	570.4	-1.29590	-1.61
226574_at	paraspeckle component 1 /// TPTE and PTEN homologous inositol lipid phosphatase pseudogene	PSPC1 /// LOC374491	592.8	-1.26174	-1.61
223455_at	trichoplein	MGC10854	408.2	-1.61216	-1.61
218940_at	chromosome 14 open reading frame 138	C14orf138	485.8	-1.44350	-1.61
244766_at	PI-3-kinase-related kinase SMG-1 /// KIAA0220-like protein /// hypothetical protein LOC440345 /// PI-3-kinase-related kinase SMG-1 pseudogene /// similar to the PI-3-kinase-related kinase SMG-1 family pseudogene 2 /// PI-3-kinase-related kinase SMG-1 - li	SMG1 /// LOC23117 /// LOC440345 /// LOC440354 /// LOC613037 /// LOC641298	400.0	-1.63231	-1.61
225338_at	Zyg-11 homolog B (C. elegans)	FLJ13456	550.8	-1.32741	-1.61
212572_at	serine/threonine kinase 38 like	STK38L	347.9	-1.77453	-1.61
219263_at	ring finger protein 128	RNF128	827.4	-0.98647	-1.61
225210_s_at	hypothetical LOC83640	MGC2560	418.5	-1.58740	-1.61
222011_s_at	t-complex 1	TCP1	401.4	-1.62892	-1.61
212050_at	WIRE protein	WIRE	632.7	-1.20508	-1.61
218605_at	transcription factor B2, mitochondrial	TFB2M	373.9	-1.70062	-1.62
226233_at	UDP-GalNAc:betaGlcNAc beta 1,3-galactosaminyltransferase, polypeptide 2	B3GALNT2	452.8	-1.51056	-1.62
201816_s_at	glioblastoma amplified sequence	GBAS	345.2	-1.78272	-1.62
218024_at	brain protein 44-like	BRP44L	483.7	-1.44749	-1.62
223042_s_at	FUN14 domain containing 2	FUNDC2	809.7	-1.00312	-1.62
209313_at	XPA binding protein 1, GTPase	XAB1	629.3	-1.20963	-1.62
202682_s_at	ubiquitin specific peptidase 4 (proto-oncogene)	USP4	668.6	-1.15806	-1.62
214281_s_at	ring finger and CHY zinc finger domain containing 1	RCHY1	387.9	-1.66321	-1.62
217717_s_at	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, beta polypeptide	YWHAB	336.3	-1.80998	-1.62
201299_s_at	MOB1, Mps One Binder kinase activator-like 1B (yeast)	MOBK1B	318.5	-1.86711	-1.62
223109_at	TruB pseudouridine (psi) synthase homolog 2 (E. coli)	TRUB2	702.3	-1.11711	-1.62
226301_at	chromosome 6 open reading frame 192	C6orf192	704.9	-1.11400	-1.62
212675_s_at	KIAA0582	KIAA0582	395.1	-1.64476	-1.62

209449_at	LSM2 homolog, U6 small nuclear RNA associated ( <i>S. cerevisiae</i> )	LSM2	423.5	-1.57576	-1.62
213573_at	Karyopherin (importin) beta 1	KPNB1	796.5	-1.01594	-1.62
204324_s_at	golgi phosphoprotein 4	GOLPH4	315.4	-1.87772	-1.62
225769_at	component of oligomeric golgi complex 6	COG6	564.6	-1.30512	-1.62
1559214_at	---	---	922.3	-0.90498	-1.62
204352_at	TNF receptor-associated factor 5	TRAF5	575.6	-1.28784	-1.62
222413_s_at	myeloid/lymphoid or mixed-lineage leukemia 3	MLL3	363.2	-1.73030	-1.62
235244_at	hypothetical LOC131076	LOC131076	360.7	-1.73736	-1.63
234464_s_at	essential meiotic endonuclease 1 homolog 1 ( <i>S. pombe</i> )	EME1	312.6	-1.88699	-1.63
217973_at	dicarbonyl/L-xylulose reductase	DCXR	439.0	-1.54063	-1.63
228248_at	TORC2-specific protein AVO3	AVO3	815.8	-0.99730	-1.63
204092_s_at	serine/threonine kinase 6	STK6	306.4	-1.90835	-1.63
220775_s_at	ubiquitin-conjugating enzyme E2-like	UEV3	723.6	-1.09260	-1.63
1553581_s_at	p18 splicing regulatory protein	P18SRP	377.5	-1.69079	-1.63
1553976_a_at	deleted in a mouse model of primary ciliary dyskinesia	RP11-529I10.4	350.9	-1.76571	-1.63
209231_s_at	dynactin 5 (p25)	DCTN5	781.7	-1.03062	-1.63
234432_at	---	---	387.7	-1.66371	-1.63
206790_s_at	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 1, 7kDa	NDUFB1	527.3	-1.36705	-1.63
233341_s_at	polymerase (RNA) I polypeptide B, 128kDa	POLR1B	614.5	-1.23032	-1.63
224851_at	cyclin-dependent kinase 6	CDK6	768.3	-1.04435	-1.63
218640_s_at	pleckstrin homology domain containing, family F (with FYVE domain) member 2	PLEKHF2	295.4	-1.94752	-1.63
227211_at	PHD finger protein 19	PHF19	871.5	-0.94695	-1.63
226661_at	cell division cycle associated 2	CDCA2	302.5	-1.92202	-1.64
222036_s_at	MCM4 minichromosome maintenance deficient 4 ( <i>S. cerevisiae</i> )	MCM4	421.8	-1.57985	-1.64
222530_s_at	McKusick-Kaufman syndrome	MKKS	301.3	-1.92640	-1.64
214022_s_at	interferon induced transmembrane protein 1 (9-27)	IFITM1	320.0	-1.86217	-1.64
213581_at	programmed cell death 2	PDCD2	479.1	-1.45651	-1.64
209406_at	BCL2-associated athanogene 2	BAG2	642.3	-1.19206	-1.64
227598_at	chromosome 7 open reading frame 29	C7orf29	317.5	-1.87052	-1.64
204976_s_at	Alport syndrome, mental retardation, midface hypoplasia and elliptocytosis chromosomal region, gene 1	AMMECR1	845.6	-0.96979	-1.64
221258_s_at	kinesin family member 18A /// kinesin family member 18A	KIF18A	384.2	-1.67299	-1.64
211988_at	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily e, member 1	SMARCE1	378.6	-1.68792	-1.64

203832_at	small nuclear ribonucleoprotein polypeptide F	SNRPF	496.8	-1.42232	-1.65
202550_s_at	VAMP (vesicle-associated membrane protein)-associated protein B and C	VAPB	559.1	-1.31379	-1.65
205664_at	KIN, antigenic determinant of recA protein homolog (mouse)	KIN	595.3	-1.25798	-1.65
225389_at	BTB (POZ) domain containing 6	BTBD6	433.3	-1.55333	-1.65
207438_s_at	RNA, U transporter 1	RNUT1	475.5	-1.46370	-1.65
202040_s_at	Jumonji, AT rich interactive domain 1A (RBBP2-like)	JARID1A	702.6	-1.11669	-1.65
226392_at	CDNA: FLJ21652 fis, clone COL08582	---	979.2	-0.86188	-1.65
212468_at	sperm associated antigen 9	SPAG9	393.1	-1.64987	-1.65
225452_at	PPAR binding protein	PPARBP	683.0	-1.14025	-1.65
221711_s_at	HSPC142 protein /// HSPC142 protein	HSPC142	720.3	-1.09636	-1.65
224754_at	Sp1 transcription factor	SP1	344.0	-1.78623	-1.65
225866_at	brix domain containing 1	BXDC1	528.0	-1.36594	-1.65
227871_at	choroideremia (Rab escort protein 1)	CHM	382.5	-1.67757	-1.65
218936_s_at	HSPC128 protein	HSPC128	488.7	-1.43777	-1.65
229025_s_at	IMP1 inner mitochondrial membrane peptidase-like (S. cerevisiae)	IMMP1L	831.4	-0.98272	-1.65
220688_s_at	chromosome 1 open reading frame 33	C1orf33	382.2	-1.67836	-1.65
209585_s_at	multiple inositol polyphosphate histidine phosphatase, 1	MINPP1	896.1	-0.92622	-1.66
201123_s_at	eukaryotic translation initiation factor 5A	EIF5A	489.7	-1.43599	-1.66
201304_at	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 5, 13kDa	NDUFA5	335.7	-1.81178	-1.66
211275_s_at	glycogenin	GYG	640.3	-1.19478	-1.66
223474_at	chromosome 14 open reading frame 4	C14orf4	823.1	-0.99045	-1.66
242059_at	Ethanolamine kinase 1	ETNK1	614.8	-1.22986	-1.66
223307_at	cell division cycle associated 3	CDCA3	604.9	-1.24401	-1.66
214499_s_at	BCL2-associated transcription factor 1	BCLAF1	472.2	-1.47048	-1.66
223154_at	mitochondrial ribosomal protein L1	MRPL1	437.5	-1.54399	-1.66
222794_x_at	PAP associated domain containing 1	PAPD1	717.8	-1.09913	-1.66
208424_s_at	cytokine induced apoptosis inhibitor 1	CIAPIN1	545.2	-1.33662	-1.66
214662_at	WD repeat domain 43	WDR43	714.3	-1.10319	-1.66
212888_at	Dicer1, Dcr-1 homolog (Drosophila)	DICER1	784.0	-1.02836	-1.66
226414_s_at	APC11 anaphase promoting complex subunit 11 homolog (yeast)	ANAPC11	522.0	-1.37637	-1.66
238755_at	Transcribed locus	---	423.0	-1.57702	-1.66
222027_at	Nuclear casein kinase and cyclin-dependent kinase substrate 1	NUCKS	444.4	-1.52865	-1.66
203049_s_at	KIAA0372	KIAA0372	419.4	-1.58540	-1.66
222740_at	ATPase family, AAA domain containing 2	ATAD2	457.1	-1.50158	-1.66
232667_at	CDNA FLJ13690 fis, clone PLACE2000097	---	425.6	-1.57096	-1.66
214001_x_at	Ribosomal protein S10	RPS10	369.1	-1.71362	-1.66

204127_at	replication factor C (activator 1) 3, 38kDa	RFC3	292.0	-1.96009	-1.66
212660_at	PHD finger protein 15	PHF15	592.3	-1.26248	-1.66
201435_s_at	eukaryotic translation initiation factor 4E	EIF4E	431.0	-1.55861	-1.66
212382_at	Transcription factor 4	TCF4	368.4	-1.71567	-1.67
221257_x_at	F-box protein 38 /// F-box protein 38	FBXO38	590.8	-1.26469	-1.67
228468_at	microtubule associated serine/threonine kinase-like	MASTL	538.4	-1.34808	-1.67
213642_at	---	---	549.4	-1.32961	-1.67
217932_at	mitochondrial ribosomal protein S7	MRPS7	418.7	-1.58694	-1.67
222906_at	feline leukemia virus subgroup C cellular receptor	FLVCR	728.7	-1.08688	-1.67
223711_s_at	thymocyte protein thy28	THY28	523.3	-1.37405	-1.67
235391_at	family with sequence similarity 92, member A1	FAM92A1	535.2	-1.35345	-1.67
204767_s_at	flap structure-specific endonuclease 1	FEN1	638.6	-1.19710	-1.67
225517_at	hypothetical protein FLJ20582	FLJ20582	598.3	-1.25354	-1.67
217853_at	Tensin 3	TENS1	659.9	-1.16911	-1.67
219237_s_at	DnaJ (Hsp40) homolog, subfamily B, member 14	DNAJB14	397.7	-1.63832	-1.67
218323_at	ras homolog gene family, member T1	RHOT1	753.9	-1.05940	-1.67
201456_s_at	BUB3 budding uninhibited by benzimidazoles 3 homolog (yeast)	BUB3	312.6	-1.88716	-1.67
226874_at	kelch-like 8 (Drosophila)	KLHL8	518.8	-1.38204	-1.67
212174_at	adenylate kinase 2	AK2	397.0	-1.63992	-1.67
225890_at	chromosome 20 open reading frame 72	C20orf72	540.3	-1.34480	-1.68
225646_at	cathepsin C	CTSC	573.6	-1.29088	-1.68
216048_s_at	Rho-related BTB domain containing 3	RHOBTB3	488.1	-1.43907	-1.68
225148_at	hypothetical protein MGC52010	MGC52010	521.4	-1.37748	-1.68
209135_at	aspartate beta-hydroxylase	ASPH	807.7	-1.00502	-1.68
208837_at	transmembrane emp24 protein transport domain containing 3	TMED3	647.7	-1.18496	-1.68
227186_s_at	mitochondrial ribosomal protein L41	MRPL41	703.2	-1.11606	-1.68
204237_at	GULP, engulfment adaptor PTB domain containing 1	GULP1	686.5	-1.13594	-1.68
212460_at	chromosome 14 open reading frame 147	C14orf147	316.0	-1.87564	-1.68
209258_s_at	chondroitin sulfate proteoglycan 6 (bamacan)	CSPG6	516.0	-1.38704	-1.68
228026_at	Hypothetical protein FLJ21168	FLJ21168	420.8	-1.58217	-1.68
214785_at	vacuolar protein sorting 13A (yeast)	VPS13A	371.5	-1.70709	-1.68
222366_at	Activity-dependent neuroprotector	ADNP	724.6	-1.09151	-1.68
218120_s_at	heme oxygenase (decycling) 2	HMOX2	314.6	-1.88027	-1.69
218212_s_at	molybdenum cofactor synthesis 2	MOCS2	1029.9	-0.82648	-1.69
211929_at	heterogeneous nuclear ribonucleoprotein A3	HNRPA3	521.5	-1.37733	-1.69
209832_s_at	DNA replication factor	CDT1	373.9	-1.70059	-1.69
212847_at	Far upstream element (FUSE) binding	FUBP1	898.4	-0.92430	-1.69

	protein 1					
222697_s_at	abhydrolase domain containing 10	ABHD10	425.8	-1.57037	-1.69	
225706_at	glucocorticoid induced transcript 1	GLCCI1	824.7	-0.98895	-1.69	
209302_at	polymerase (RNA) II (DNA directed) polypeptide H	POLR2H	566.8	-1.30159	-1.69	
228737_at	chromosome 20 open reading frame 100	C20orf100	653.1	-1.17786	-1.69	
212585_at	oxysterol binding protein-like 8	OSBPL8	320.6	-1.86022	-1.69	
203362_s_at	MAD2 mitotic arrest deficient-like 1 (yeast)	MAD2L1	463.0	-1.48920	-1.69	
200907_s_at	palladin	KIAA0992	366.2	-1.72191	-1.69	
226350_at	choroideremia-like (Rab escort protein 2)	CHML	483.9	-1.44720	-1.69	
1553575_at	---	---	411.1	-1.60528	-1.69	
214414_x_at	hemoglobin, alpha 2 /// hemoglobin, alpha 2	HBA2	502.0	-1.41261	-1.69	
224649_x_at	chromosome 10 open reading frame 9	C10orf9	750.7	-1.06286	-1.69	
201740_at	NADH dehydrogenase (ubiquinone) Fe-S protein 3, 30kDa (NADH-coenzyme Q reductase)	NDUFS3	602.7	-1.24719	-1.70	
226898_s_at	Splicing factor proline/glutamine-rich (polypyrimidine tract binding protein associated)	SFPQ	297.1	-1.94138	-1.70	
214835_s_at	succinate-CoA ligase, GDP-forming, beta subunit	SUCLG2	805.6	-1.00708	-1.70	
212402_at	KIAA0853	KIAA0853	732.0	-1.08324	-1.70	
228980_at	ring finger and FYVE-like domain containing 1	RFFL	631.1	-1.20719	-1.70	
238040_at	Pogo transposable element with ZNF domain	POGZ	454.0	-1.50801	-1.70	
225581_s_at	mitochondrial ribosomal protein L50	MRPL50	643.4	-1.19062	-1.70	
221452_s_at	transmembrane protein 14B /// transmembrane protein 14B	TMEM14B	693.7	-1.12722	-1.70	
213572_s_at	serpin peptidase inhibitor, clade B (ovalbumin), member 1	SERPINB1	443.4	-1.53086	-1.70	
224840_at	FK506 binding protein 5	FKBP5	404.9	-1.62023	-1.70	
205733_at	Bloom syndrome	BLM	295.1	-1.94856	-1.70	
218195_at	chromosome 6 open reading frame 211	C6orf211	623.2	-1.21812	-1.70	
221561_at	sterol O-acyltransferase (acyl-Coenzyme A: cholesterol acyltransferase) 1	SOAT1	547.5	-1.33277	-1.70	
214697_s_at	ROD1 regulator of differentiation 1 (S. pombe)	ROD1	324.7	-1.84690	-1.70	
242652_at	Erythrocyte membrane protein band 4.1- like 2	EPB41L2	468.4	-1.47801	-1.70	
225861_at	hypothetical protein MGC15416	MGC15416	821.6	-0.99186	-1.70	
213325_at	poliovirus receptor-related 3	PVRL3	632.1	-1.20583	-1.70	
200864_s_at	RAB11A, member RAS oncogene family	RAB11A	603.9	-1.24545	-1.71	
227012_at	mitochondrial carrier family protein	MCFP	915.8	-0.91014	-1.71	
212364_at	myosin IB	MYO1B	674.7	-1.15046	-1.71	

222808_at	glycosyltransferase 28 domain containing 1	GLT28D1	833.1	-0.98112	-1.71
233268_s_at	churchill domain containing 1	CHURC1	530.3	-1.36189	-1.71
227476_at	Lysophosphatidylglycerol acyltransferase 1	LPGAT1	319.6	-1.86340	-1.71
219372_at	carnitine deficiency-associated, expressed in ventricle 1	CDV1	642.6	-1.19166	-1.71
221884_at	ecotropic viral integration site 1	EVI1	428.8	-1.56367	-1.71
203449_s_at	telomeric repeat binding factor (NIMA-interacting) 1	TERF1	415.9	-1.59372	-1.71
210596_at	---	---	355.4	-1.75244	-1.71
209172_s_at	centromere protein F, 350/400ka (mitosin) /// centromere protein F, 350/400ka (mitosin)	CENPF	292.7	-1.95757	-1.71
227689_at	zinc finger protein 227	ZNF227	560.9	-1.31104	-1.71
202760_s_at	A kinase (PRKA) anchor protein 2 /// PALM2-AKAP2 protein	AKAP2 /// PALM2-AKAP2	494.6	-1.42659	-1.71
223592_s_at	ring finger protein 135	RNF135	733.4	-1.08168	-1.71
225352_at	translocation protein 1	TLOC1	431.7	-1.55696	-1.72
209787_s_at	high mobility group nucleosomal binding domain 4	HMGN4	437.1	-1.54485	-1.72
33778_at	TBC1 domain family, member 22A	TBC1D22A	670.8	-1.15529	-1.72
201083_s_at	BCL2-associated transcription factor 1	BCLAF1	752.0	-1.06150	-1.72
227278_at	Transcribed locus, moderately similar to XP_512541.1 PREDICTED: similar to hypothetical protein [Pan troglodytes]	---	401.8	-1.62803	-1.72
227776_at	Transcribed locus	---	685.1	-1.13758	-1.72
212603_at	mitochondrial ribosomal protein S31	MRPS31	353.9	-1.75682	-1.72
202599_s_at	nuclear receptor interacting protein 1	NRIP1	389.3	-1.65955	-1.72
1557411_s_at	similar to solute carrier family 25 , member 16	LOC203427	348.1	-1.77401	-1.72
225295_at	solute carrier family 39 (zinc transporter), member 10	SLC39A10	512.7	-1.39309	-1.72
212268_at	serpin peptidase inhibitor, clade B (ovalbumin), member 1	SERPINB1	569.0	-1.29808	-1.72
204709_s_at	kinesin family member 23	KIF23	353.3	-1.75879	-1.72
235521_at	homeo box A3	HOXA3	390.8	-1.65572	-1.72
219819_s_at	mitochondrial ribosomal protein S28	MRPS28	429.6	-1.56184	-1.72
204521_at	protein predicted by clone 23733	HSU79274	684.3	-1.13862	-1.73
204897_at	prostaglandin E receptor 4 (subtype EP4)	PTGER4	836.4	-0.97810	-1.73
230618_s_at	BAT2 domain containing 1	XTP2	538.1	-1.34864	-1.73
209513_s_at	hydroxysteroid dehydrogenase like 2	HSDL2	531.8	-1.35934	-1.73
209773_s_at	ribonucleotide reductase M2 polypeptide	RRM2	388.1	-1.66275	-1.73
206430_at	caudal type homeo box transcription factor 1	CDX1	814.9	-0.99813	-1.73
224684_at	Sorting nexin 12	SNX12	394.6	-1.64614	-1.73
222762_x_at	LIM domains containing 1	LIMD1	379.3	-1.68608	-1.73

213109_at	TRAF2 and NCK interacting kinase	TNIK	435.2	-1.54917	-1.73
203448_s_at	telomeric repeat binding factor (NIMA-interacting) 1	TERF1	487.8	-1.43953	-1.73
212205_at	H2A histone family, member V	H2AFV	712.5	-1.10522	-1.73
203060_s_at	3'-phosphoadenosine 5'-phosphosulfate synthase 2	PAPSS2	421.1	-1.58150	-1.73
222613_at	chromosome 12 open reading frame 4	C12orf4	851.9	-0.96416	-1.74
223021_x_at	chromosome 6 open reading frame 55	C6orf55	594.8	-1.25873	-1.74
225415_at	deltex 3-like (Drosophila)	DTX3L	711.3	-1.10662	-1.74
216609_at	Thioredoxin	TXN	340.8	-1.79614	-1.74
209030_s_at	immunoglobulin superfamily, member 4	IGSF4	519.5	-1.38084	-1.74
208955_at	dUTP pyrophosphatase	DUT	612.4	-1.23324	-1.74
228690_s_at	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 11, 14.7kDa	NDUFA11	682.8	-1.14041	-1.74
225153_at	G elongation factor, mitochondrial 1	GFM1	295.0	-1.94905	-1.75
202157_s_at	CUG triplet repeat, RNA binding protein 2	CUGBP2	729.0	-1.08655	-1.75
204641_at	NIMA (never in mitosis gene a)-related kinase 2	NEK2	416.3	-1.59267	-1.75
210559_s_at	cell division cycle 2, G1 to S and G2 to M	CDC2	601.2	-1.24939	-1.75
226117_at	TRAF-interacting protein with a forkhead-associated domain	TIFA	946.7	-0.88607	-1.75
203306_s_at	solute carrier family 35 (CMP-sialic acid transporter), member A1	SLC35A1	869.4	-0.94880	-1.75
218286_s_at	ring finger protein 7	RNF7	537.3	-1.34998	-1.75
204999_s_at	activating transcription factor 5	ATF5	595.6	-1.25756	-1.75
224576_at	endoplasmic reticulum-golgi intermediate compartment 32 kDa protein	KIAA1181	392.2	-1.65223	-1.75
202298_at	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 1, 7.5kDa	NDUFA1	929.6	-0.89921	-1.75
224644_at	CDNA clone IMAGE:5278517	---	321.4	-1.85760	-1.75
226194_at	chromosome 13 open reading frame 8	C13orf8	446.9	-1.52330	-1.75
204370_at	ATP/GTP-binding protein	HEAB	459.3	-1.49693	-1.75
234675_x_at	CDNA: FLJ23566 fis, clone LNG10880	---	764.6	-1.04819	-1.76
225304_s_at	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 11, 14.7kDa	NDUFA11	609.8	-1.23688	-1.76
212378_at	phosphoribosylglycinamide formyltransferase, phosphoribosylglycinamide synthetase, phosphoribosylaminoimidazole synthetase	GART	376.9	-1.69240	-1.76
220011_at	chromosome 1 open reading frame 135	C1orf135	445.4	-1.52645	-1.76
226157_at	Transcription factor Dp-2 (E2F dimerization partner 2)	TFDP2	423.6	-1.57553	-1.76
226962_at	zinc finger and BTB domain containing 41	ZBTB41	1056.2	-0.80917	-1.76
225358_at	DnaJ (Hsp40) homolog, subfamily C, member 19	DNAJC19	806.6	-1.00611	-1.76
241397_at	Ets homologous factor	EHF	937.5	-0.89311	-1.77

204388_s_at	monoamine oxidase A	MAOA	464.3	-1.48652	-1.77
203837_at	mitogen-activated protein kinase kinase kinase 5	MAP3K5	544.1	-1.33847	-1.77
224711_at	YY1 transcription factor	YY1	505.4	-1.40644	-1.77
226426_at	Activity-dependent neuroprotector	ADNP	757.4	-1.05571	-1.77
225253_s_at	methyltransferase like 2 /// hypothetical protein FLJ12760	METTL2 /// FLJ12760	475.0	-1.46474	-1.77
213494_s_at	YY1 transcription factor	YY1	657.2	-1.17262	-1.77
221786_at	chromosome 6 open reading frame 120	C6orf120	430.9	-1.55875	-1.77
227420_at	tumor necrosis factor, alpha-induced protein 8-like 1	TNFAIP8L1	570.6	-1.29559	-1.77
202124_s_at	amyotrophic lateral sclerosis 2 (juvenile) chromosome region, candidate 3	ALS2CR3	772.4	-1.04016	-1.77
243543_at	Sterol-C4-methyl oxidase-like	SC4MOL	882.6	-0.93752	-1.77
226834_at	Adipocyte-specific adhesion molecule	ASAM	778.4	-1.03405	-1.77
1557965_at	MTERF domain containing 2	MTERFD2	681.7	-1.14174	-1.78
221488_s_at	chromosome 6 open reading frame 82	C6orf82	493.2	-1.42927	-1.78
221531_at	WD repeat domain 61	WDR61	815.4	-0.99767	-1.78
200078_s_at	ATPase, H <sup>+</sup> transporting, lysosomal 21kDa, V0 subunit c" /// ATPase, H <sup>+</sup> transporting, lysosomal 21kDa, V0 subunit c"	ATP6V0B	683.5	-1.13960	-1.78
218264_at	BRCA2 and CDKN1A interacting protein	BCCIP	532.4	-1.35831	-1.78
209250_at	degenerative spermatocyte homolog 1, lipid desaturase (Drosophila)	DEGS1	985.8	-0.85709	-1.78
226686_at	similar to RIKEN cDNA 1500009M05 gene	LOC493856	854.8	-0.96155	-1.78
218334_at	Ngg1 interacting factor 3 like 1 binding protein 1	NIF3L1BP1	422.9	-1.57723	-1.78
212926_at	SMC5 structural maintenance of chromosomes 5-like 1 (yeast)	SMC5L1	818.2	-0.99500	-1.78
208673_s_at	splicing factor, arginine/serine-rich 3	SFRS3	488.2	-1.43874	-1.78
222453_at	cytochrome b reductase 1	CYBRD1	915.7	-0.91029	-1.78
204508_s_at	carbonic anhydrase XII	CA12	547.1	-1.33353	-1.78
214041_x_at	Ribosomal protein L37a	RPL37A	684.4	-1.13851	-1.78
233849_s_at	Rho GTPase activating protein 5	ARHGAP5	405.4	-1.61909	-1.78
221823_at	hypothetical gene supported by AF038182; BC009203	LOC90355	497.3	-1.42153	-1.79
212900_at	SEC24 related gene family, member A (S. cerevisiae)	SEC24A	694.0	-1.12688	-1.79
228956_at	UDP glycosyltransferase 8 (UDP-galactose ceramide galactosyltransferase)	UGT8	613.7	-1.23137	-1.79
227484_at	CDNA FLJ41690 fis, clone HCASM2009405	---	487.6	-1.43995	-1.79
229467_at	Poly(rC) binding protein 2	PCBP2	900.5	-0.92262	-1.79

225198_at	VAMP (vesicle-associated membrane protein)-associated protein A, 33kDa	VAPA	866.1	-0.95163	-1.79
229026_at	CDC42 small effector 2	CDC42SE2	399.8	-1.63304	-1.79
1555758_a_at	cyclin-dependent kinase inhibitor 3 (CDK2-associated dual specificity phosphatase)	CDKN3	741.2	-1.07306	-1.79
235093_at	Transcribed locus, moderately similar to XP_524454.1 PREDICTED: hypothetical protein XP_524454 [Pan troglodytes]	---	426.7	-1.56844	-1.79
201070_x_at	splicing factor 3b, subunit 1, 155kDa	SF3B1	465.0	-1.48501	-1.79
221493_at	TSPY-like 1	TSPYL1	491.9	-1.43172	-1.80
201197_at	adenosylmethionine decarboxylase 1	AMD1	452.2	-1.51190	-1.80
201580_s_at	thioredoxin domain containing 13	TXNDC13	546.8	-1.33390	-1.80
222424_s_at	nuclear casein kinase and cyclin-dependent kinase substrate 1	NUCKS1	285.5	-1.98463	-1.80
217811_at	selenoprotein T	SELT	283.2	-1.99315	-1.80
219023_at	chromosome 4 open reading frame 16	C4orf16	700.3	-1.11945	-1.80
226170_at	---	---	525.0	-1.37109	-1.80
201340_s_at	ectodermal-neural cortex (with BTB-like domain)	ENC1	457.2	-1.50128	-1.80
225856_at	CDNA FLJ39000 fis, clone NT2RI2022468	---	415.5	-1.59472	-1.80
225647_s_at	cathepsin C	CTSC	382.1	-1.67854	-1.80
212648_at	DEAH (Asp-Glu-Ala-His) box polypeptide 29	DHX29	333.6	-1.81849	-1.80
203420_at	family with sequence similarity 8, member A1	FAM8A1	680.6	-1.14315	-1.80
222958_s_at	DEP domain containing 1	DEPDC1	343.6	-1.78743	-1.81
225834_at	family with sequence similarity 72, member A	FAM72A	615.5	-1.22888	-1.81
1564637_a_at	hypothetical protein FLJ38426	FLJ38426	510.9	-1.39626	-1.81
223331_s_at	DEAD (Asp-Glu-Ala-Asp) box polypeptide 20	DDX20	505.9	-1.40543	-1.81
242140_at	similar to envelope protein	LOC113386	462.6	-1.49003	-1.81
227650_at	heat shock 70kDa protein 14	HSPA14	583.3	-1.27603	-1.81
208848_at	alcohol dehydrogenase 5 (class III), chi polypeptide	ADH5	723.0	-1.09330	-1.81
223701_s_at	ubiquitin specific peptidase 47	USP47	541.7	-1.34256	-1.81
213995_at	ATP synthase, H <sup>+</sup> transporting, mitochondrial F0 complex, subunit s (factor B)	ATP5S	1004.7	-0.84375	-1.81
226748_at	LysM, putative peptidoglycan-binding, domain containing 2	LYSMD2	426.6	-1.56853	-1.81
225256_at	CDNA FLJ41369 fis, clone BRCAN2006117	---	295.5	-1.94739	-1.81
226974_at	Neural precursor cell expressed, developmentally down-regulated 4-like	NEDD4L	601.4	-1.24912	-1.81

215338_s_at	natural killer-tumor recognition sequence	NKTR	605.3	-1.24335	-1.82
211318_s_at	RAE1 RNA export 1 homolog (S. pombe)	RAE1	347.6	-1.77549	-1.82
229742_at	hypothetical LOC145853	LOC145853	778.0	-1.03436	-1.82
225470_at	nucleoporin 35kDa	NUP35	850.0	-0.96584	-1.82
222409_at	coronin, actin binding protein, 1C	CORO1C	655.3	-1.17505	-1.82
223342_at	ribonucleotide reductase M2 B (TP53 inducible)	RRM2B	719.3	-1.09741	-1.82
225431_x_at	aminoacylase 1-like 2	ACY1L2	386.0	-1.66816	-1.82
203983_at	translin-associated factor X	TSNAX	1122.8	-0.76801	-1.83
218163_at	malignant T cell amplified sequence 1	MCTS1	365.2	-1.72446	-1.83
228496_s_at	Cysteine rich transmembrane BMP regulator 1 (chordin-like)	CRIM1	295.5	-1.94745	-1.83
228520_s_at	Amyloid beta (A4) precursor-like protein 2	APLP2	579.4	-1.28190	-1.83
221802_s_at	KIAA1598	KIAA1598	787.8	-1.02455	-1.83
209684_at	Ras and Rab interactor 2	RIN2	501.1	-1.41437	-1.83
228853_at	---	---	722.8	-1.09352	-1.83
202432_at	protein phosphatase 3 (formerly 2B), catalytic subunit, beta isoform (calcineurin A beta)	PPP3CB	710.1	-1.10796	-1.83
203974_at	haloacid dehalogenase-like hydrolase domain containing 1A	HDHD1A	479.0	-1.45687	-1.83
203836_s_at	mitogen-activated protein kinase kinase 5	MAP3K5	521.4	-1.37753	-1.83
243904_at	CDNA clone IMAGE:5287121	---	527.9	-1.36602	-1.84
203606_at	NADH dehydrogenase (ubiquinone) Fe-S protein 6, 13kDa (NADH-coenzyme Q reductase)	NDUFS6	663.7	-1.16425	-1.84
218042_at	COP9 constitutive photomorphogenic homolog subunit 4 (Arabidopsis)	COPS4	539.1	-1.34685	-1.84
212605_s_at	Nudix (nucleoside diphosphate linked moiety X)-type motif 3	NUDT3	472.1	-1.47049	-1.84
227239_at	down-regulated by Ctnnb1, a	DRCTNNB1A	655.4	-1.17485	-1.84
209748_at	spastin	SPAST	398.6	-1.63607	-1.84
244669_at	Chromosome 6 open reading frame 160	chromosome 6 open reading frame 160	718.0	-1.09893	-1.84
213813_x_at	Ferritin, light polypeptide	FTL	585.3	-1.27300	-1.84
202634_at	polymerase (RNA) II (DNA directed) polypeptide K, 7.0kDa	POLR2K	589.9	-1.26613	-1.84
218564_at	ring finger and WD repeat domain 3	RFWD3	739.7	-1.07472	-1.84
224446_at	hypothetical protein MGC14817 /// hypothetical protein MGC14817	MGC14817	564.9	-1.30462	-1.85
1555826_at	Effector cell peptidase receptor 1	BIRC5	549.3	-1.32975	-1.85
228009_x_at	zinc ribbon domain containing, 1	ZNRD1	399.9	-1.63272	-1.85

1555225_at	chromosome 1 open reading frame 43	C1orf43	521.7	-1.37693	-1.85
227636_at	THAP domain containing 5	THAP5	561.1	-1.31074	-1.85
213132_s_at	malonyl-CoA:acyl carrier protein transacylase, mitochondrial	MT	469.6	-1.47564	-1.85
205133_s_at	heat shock 10kDa protein 1 (chaperonin 10)	HSPE1	924.5	-0.90325	-1.85
200096_s_at	ATPase, H <sup>+</sup> transporting, lysosomal 9kDa, V0 subunit e /// ATPase, H <sup>+</sup> transporting, lysosomal 9kDa, V0 subunit e	ATP6V0E	329.0	-1.83286	-1.85
218568_at	multiple substrate lipid kinase	MULK	524.0	-1.37285	-1.85
223225_s_at	SEH1-like (S. cerevisiae)	SEH1L	484.7	-1.44555	-1.85
202846_s_at	phosphatidylinositol glycan, class C	PIGC	454.0	-1.50801	-1.85
202798_at	SEC24 related gene family, member B (S. cerevisiae)	SEC24B	718.2	-1.09873	-1.85
212917_x_at	RecQ protein-like (DNA helicase Q1-like)	RECQL	790.0	-1.02232	-1.85
229043_at	PAP associated domain containing 5	PAPD5	537.3	-1.34987	-1.86
219037_at	CGI-115 protein	CGI-115	399.5	-1.63379	-1.86
227249_at	Myosin, heavy polypeptide 11, smooth muscle	MYH11	732.0	-1.08323	-1.86
214711_at	hypothetical protein LOC283459	15E1.2	736.5	-1.07825	-1.86
228092_at	cAMP responsive element modulator	CREM	876.1	-0.94303	-1.86
213239_at	chromosome 13 open reading frame 24	C13orf24	652.3	-1.17896	-1.86
201890_at	ribonucleotide reductase M2 polypeptide	RRM2	580.4	-1.28046	-1.86
226300_at	mediator of RNA polymerase II transcription, subunit 19 homolog (yeast)	MED19	608.5	-1.23877	-1.86
212513_s_at	ubiquitin specific peptidase 33	USP33	597.6	-1.25471	-1.86
1570561_at	---	---	370.0	-1.71134	-1.87
224206_x_at	myoneurin	MYNN	832.9	-0.98132	-1.87
226861_at	ankyrin repeat and SOCS box-containing 8	ASB8	1206.8	-0.72118	-1.87
226165_at	chromosome 8 open reading frame 59	C8orf59	698.5	-1.12160	-1.87
218527_at	aprataxin	APTX	961.9	-0.87456	-1.87
211745_x_at	hemoglobin, alpha 1 /// hemoglobin, alpha 1 /// hemoglobin, alpha 2 /// hemoglobin, alpha 2	HBA1 /// HBA2	688.4	-1.13368	-1.87
206848_at	homeo box A7	HOXA7	302.0	-1.92375	-1.87
202172_at	zinc finger protein 161	ZNF161	970.5	-0.86820	-1.87
235425_at	shugoshin-like 2 (S. pombe)	SGOL2	600.0	-1.25108	-1.87
200623_s_at	calmodulin 3 (phosphorylase kinase, delta)	CALM3	516.6	-1.38607	-1.87
228969_at	anterior gradient 2 homolog (Xenopus laevis)	AGR2	856.5	-0.96008	-1.87
215884_s_at	ubiquilin 2	UBQLN2	749.2	-1.06444	-1.87
227174_at	WD repeat domain 72	WDR72	939.0	-0.89194	-1.87
219802_at	hypothetical protein FLJ22028	FLJ22028	1078.0	-0.79526	-1.87
202623_at	chromosome 14 open reading frame 11	C14orf11	706.7	-1.11195	-1.87
222824_at	Sec61 alpha 2 subunit (S. cerevisiae)	SEC61A2	599.7	-1.25162	-1.87

202739_s_at	phosphorylase kinase, beta	PHKB	753.0	-1.06039	-1.88
242665_at	formin-like 2	FMNL2	538.6	-1.34775	-1.88
223065_s_at	STARD3 N-terminal like	STARD3NL	496.4	-1.42307	-1.88
224159_x_at	tripartite motif-containing 4	TRIM4	531.3	-1.36022	-1.88
203581_at	RAB4A, member RAS oncogene family	RAB4A	444.0	-1.52961	-1.88
213971_s_at	suppressor of zeste 12 homolog (Drosophila)	SUZ12	558.9	-1.31423	-1.88
221478_at	BCL2/adenovirus E1B 19kDa interacting protein 3-like /// BCL2/adenovirus E1B 19kDa interacting protein 3-like	BNIP3L	584.2	-1.27465	-1.88
225195_at	zinc finger, CSL-type containing 2	ZCSL2	309.4	-1.89815	-1.88
218617_at	tRNA isopentenyltransferase 1	TRIT1	954.5	-0.88015	-1.89
202900_s_at	nucleoporin 88kDa	NUP88	297.2	-1.94118	-1.89
201202_at	proliferating cell nuclear antigen	PCNA	404.4	-1.62158	-1.89
234295_at	debranching enzyme homolog 1 (S. cerevisiae)	DBR1	821.1	-0.99229	-1.89
201990_s_at	cAMP responsive element binding protein-like 2	CREBL2	635.6	-1.20113	-1.89
203132_at	retinoblastoma 1 (including osteosarcoma)	RB1	542.8	-1.34068	-1.89
228992_at	Mediator of RNA polymerase II transcription, subunit 28 homolog (yeast)	MED28	1039.3	-0.82019	-1.89
200653_s_at	calmodulin 1 (phosphorylase kinase, delta)	CALM1	473.9	-1.46697	-1.89
225467_s_at	retinol dehydrogenase 13 (all-trans and 9-cis)	RDH13	417.5	-1.58975	-1.89
239050_s_at	CDNA FLJ13202 fis, clone NT2RP3004503	---	407.1	-1.61499	-1.90
219221_at	zinc finger and BTB domain containing 38	ZBTB38	1063.0	-0.80476	-1.90
228729_at	cyclin B1	CCNB1	615.1	-1.22938	-1.90
211754_s_at	solute carrier family 25 (mitochondrial carrier; peroxisomal membrane protein, 34kDa), member 17 /// solute carrier family 25 (mitochondrial carrier; peroxisomal membrane protein, 34kDa), member 17	SLC25A17	688.7	-1.13331	-1.90
203284_s_at	heparan sulfate 2-O-sulfotransferase 1	HS2ST1	304.0	-1.91684	-1.90
201477_s_at	ribonucleotide reductase M1 polypeptide	RRM1	616.5	-1.22740	-1.90
208804_s_at	splicing factor, arginine/serine-rich 6	SFRS6	462.4	-1.49042	-1.90
213294_at	Hypothetical protein FLJ38348	FLJ38348	659.3	-1.16984	-1.90
203216_s_at	myosin VI	MYO6	444.4	-1.52878	-1.90
208898_at	ATPase, H <sup>+</sup> transporting, lysosomal 34kDa, V1 subunit D	ATP6V1D	552.3	-1.32496	-1.90
1563321_s_at	myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila); translocated to, 10	MLLT10	1024.9	-0.82983	-1.91
225580_at	mitochondrial ribosomal protein L50	MRPL50	619.7	-1.22300	-1.91
212751_at	ubiquitin-conjugating enzyme E2N (UBC13 homolog, yeast)	UBE2N	727.8	-1.08787	-1.91
226025_at	ankyrin repeat domain 28	ANKRD28	684.4	-1.13849	-1.91

224735_at	cytochrome b, ascorbate dependent 3	CYBASC3	519.2	-1.38144	-1.91
212437_at	centromere protein B, 80kDa	CENPB	415.7	-1.59417	-1.91
211711_s_at	phosphatase and tensin homolog (mutated in multiple advanced cancers 1) /// phosphatase and tensin homolog (mutated in multiple advanced cancers 1)	PTEN	789.2	-1.02317	-1.91
212652_s_at	sorting nexin 4	SNX4	888.2	-0.93279	-1.91
224786_at	short coiled-coil protein	SCOC	969.0	-0.86933	-1.91
217813_s_at	spindlin	SPIN	1024.5	-0.83016	-1.92
225644_at	hypothetical protein FLJ33814	FLJ33814	984.2	-0.85827	-1.92
213812_s_at	calcium/calmodulin-dependent protein kinase kinase 2, beta	CAMKK2	1160.4	-0.74638	-1.92
223361_at	chromosome 6 open reading frame 115	C6orf115	424.4	-1.57379	-1.92
225106_s_at	hypothetical protein FLJ10826	FLJ10826	333.0	-1.82017	-1.92
213929_at	Homo sapiens, Similar to likely ortholog of yeast ARV1, clone IMAGE:4733238, mRNA	---	973.0	-0.86637	-1.92
212867_at	Nuclear receptor coactivator 2 /// Nuclear receptor coactivator 2	NCOA2	890.6	-0.93074	-1.92
222988_s_at	transmembrane protein 9	TMEM9	957.3	-0.87805	-1.92
226853_at	BMP2 inducible kinase	BMP2K	568.1	-1.29952	-1.92
204386_s_at	mitochondrial ribosomal protein 63	MRP63	688.3	-1.13376	-1.92
212589_at	Sterol carrier protein 2	SCP2	352.0	-1.76246	-1.92
213702_x_at	N-acylsphingosine amidohydrolase (acid ceramidase) 1	ASAHI	939.5	-0.89155	-1.92
225712_at	gem (nuclear organelle) associated protein 5	GEMIN5	704.7	-1.11431	-1.92
204235_s_at	GULP, engulfment adaptor PTB domain containing 1	GULP1	604.0	-1.24525	-1.92
213372_at	progesterin and adipoQ receptor family member III	PAQR3	731.9	-1.08329	-1.92
225953_at	hypothetical protein FLJ10656	P15RS	453.9	-1.50833	-1.93
225290_at	MRNA; cDNA DKFZp566C034 (from clone DKFZp566C034)	---	367.4	-1.71841	-1.93
231855_at	KIAA1524	KIAA1524	794.9	-1.01751	-1.93
218668_s_at	RAP2C, member of RAS oncogene family	RAP2C	398.2	-1.63693	-1.93
231252_at	hypothetical protein FLJ23861	FLJ23861	704.1	-1.11495	-1.93
209507_at	replication protein A3, 14kDa	RPA3	718.1	-1.09879	-1.93
226223_at	PRKC, apoptosis, WT1, regulator	PAWR	461.9	-1.49152	-1.93
235918_x_at	---	---	777.4	-1.03504	-1.93
222637_at	COMM domain containing 10	COMMD10	1428.4	-0.61878	-1.93
213623_at	kinesin family member 3A	KIF3A	852.8	-0.96335	-1.93
212709_at	nucleoporin 160kDa	NUP160	524.0	-1.37295	-1.93
200984_s_at	CD59 antigen p18-20 (antigen identified by monoclonal antibodies 16.3A5, EJ16, EJ30, EL32 and G344)	CD59	825.0	-0.98866	-1.93
213998_s_at	DEAD (Asp-Glu-Ala-Asp) box	DDX17	832.4	-0.98182	-1.94

	polypeptide 17					
219751_at	hypothetical protein FLJ21148	FLJ21148	885.2	-0.93529	-1.94	
217506_at	Hypothetical gene supported by BC041875; BX648984	---	870.2	-0.94812	-1.94	
224648_at	GC-rich promoter binding protein 1	GPBP1	899.5	-0.92338	-1.94	
229846_s_at	mitogen-activated protein kinase associated protein 1	MAPKAP1	640.2	-1.19485	-1.95	
208666_s_at	suppression of tumorigenicity 13 (colon carcinoma) (Hsp70 interacting protein)	ST13	709.6	-1.10853	-1.95	
218049_s_at	mitochondrial ribosomal protein L13	MRPL13	359.3	-1.74142	-1.95	
213574_s_at	Karyopherin (importin) beta 1	KPNB1	820.7	-0.99265	-1.95	
32099_at	scaffold attachment factor B2	SAFB2	521.3	-1.37771	-1.95	
217986_s_at	bromodomain adjacent to zinc finger domain, 1A	BAZ1A	682.1	-1.14129	-1.95	
208358_s_at	UDP glycosyltransferase 8 (UDP-galactose ceramide galactosyltransferase)	UGT8	652.8	-1.17832	-1.95	
236488_s_at	CDNA FLJ36309 fis, clone THYMU2004986	---	705.6	-1.11316	-1.96	
227198_at	AF4/FMR2 family, member 3	AFF3	301.3	-1.92629	-1.96	
209300_s_at	NECAP endocytosis associated 1	NECAP1	1007.8	-0.84160	-1.96	
235451_at	SMAD, mothers against DPP homolog 5 (Drosophila)	SMAD5	898.9	-0.92393	-1.96	
221230_s_at	AT rich interactive domain 4B (RBP1-like) /// AT rich interactive domain 4B (RBP1- like)	ARID4B	386.8	-1.66621	-1.96	
209537_at	exostoses (multiple)-like 2	EXTL2	1215.7	-0.71653	-1.96	
223482_at	transmembrane protein induced by tumor necrosis factor alpha	TMPIT	877.7	-0.94165	-1.96	
203214_x_at	cell division cycle 2, G1 to S and G2 to M	CDC2	706.5	-1.11216	-1.96	
205070_at	inhibitor of growth family, member 3	ING3	1113.9	-0.77333	-1.96	
205052_at	AU RNA binding protein/enoyl-Coenzyme A hydratase	AUH	456.4	-1.50297	-1.96	
230165_at	shugoshin-like 2 (S. pombe)	SGOL2	841.0	-0.97390	-1.96	
203970_s_at	peroxisomal biogenesis factor 3	PEX3	793.2	-1.01918	-1.96	
229257_at	KIAA1856 protein	KIAA1856	717.9	-1.09900	-1.96	
243683_at	Mortality factor 4 like 2	MORF4L2	390.3	-1.65707	-1.96	
213005_s_at	ankyrin repeat domain 15	ANKRD15	723.5	-1.09267	-1.96	
218116_at	chromosome 9 open reading frame 78 /// chromosome 9 open reading frame 78	C9orf78	1172.2	-0.73986	-1.97	
208712_at	cyclin D1	CCND1	1026.0	-0.82910	-1.97	
201362_at	influenza virus NS1A binding protein	IVNS1ABP	1057.3	-0.80846	-1.97	
202299_s_at	hepatitis B virus x interacting protein	HBXIP	1109.4	-0.77599	-1.97	
211733_x_at	sterol carrier protein 2 /// sterol carrier protein 2	SCP2	566.2	-1.30253	-1.97	
227021_at	amine oxidase (flavin containing) domain 1	AOF1	404.5	-1.62116	-1.97	
227139_s_at	Hermansky-Pudlak syndrome 3	HPS3	620.9	-1.22120	-1.97	

218202_x_at	mitochondrial ribosomal protein L44	MRPL44	682.5	-1.14084	-1.98
225884_s_at	zinc finger protein 336	ZNF336	523.9	-1.37315	-1.98
225227_at	SKI-like	SKIL	631.3	-1.20686	-1.98
204822_at	TTK protein kinase	TTK	679.2	-1.14487	-1.98
218025_s_at	peroxisomal D3,D2-enoyl-CoA isomerase	PECI	886.3	-0.93440	-1.98
205644_s_at	small nuclear ribonucleoprotein polypeptide G	SNRPG	797.9	-1.01455	-1.98
218518_at	chromosome 5 open reading frame 5	C5orf5	1402.6	-0.62940	-1.98
203275_at	interferon regulatory factor 2	IRF2	1014.7	-0.83682	-1.98
212060_at	U2-associated SR140 protein	SR140	678.0	-1.14638	-1.98
203032_s_at	fumarate hydratase	FH	737.4	-1.07720	-1.99
209482_at	processing of precursor 7, ribonuclease P subunit ( <i>S. cerevisiae</i> )	POP7	616.3	-1.22771	-1.99
218493_at	chromosome 16 open reading frame 33	C16orf33	370.9	-1.70868	-1.99
201243_s_at	ATPase, Na <sup>+</sup> /K <sup>+</sup> transporting, beta 1 polypeptide	ATP1B1	671.5	-1.15446	-1.99
225889_at	AE binding protein 2	AEBP2	670.3	-1.15596	-1.99
217987_at	putative protein product of Nbla00058	NBLA00058	473.1	-1.46850	-1.99
218663_at	chromosome condensation protein G	HCAP-G	795.6	-1.01683	-1.99
224844_at	KIAA1458 protein	KIAA1458	990.2	-0.85400	-2.00
235103_at	Mannosidase, alpha, class 2A, member 1	MAN2A1	709.0	-1.10927	-2.00
219105_x_at	origin recognition complex, subunit 6 homolog-like ( <i>yeast</i> )	ORC6L	836.0	-0.97850	-2.00
230180_at	DEAD (Asp-Glu-Ala-Asp) box polypeptide 17	DDX17	755.3	-1.05799	-2.00
227960_s_at	fumarylacetoacetate hydrolase domain containing 1	FAHD1	429.2	-1.56273	-2.00
213077_at	YTH domain containing 2	YTHDC2	591.3	-1.26398	-2.00
37462_i_at	splicing factor 3a, subunit 2, 66kDa	SF3A2	890.6	-0.93077	-2.00
225219_at	SMAD, mothers against DPP homolog 5 ( <i>Drosophila</i> )	SMAD5	531.5	-1.35977	-2.00
223451_s_at	chemokine-like factor	CKLF	1181.0	-0.73500	-2.00
228082_at	adipocyte-specific adhesion molecule	ASAM	597.7	-1.25452	-2.01
218106_s_at	mitochondrial ribosomal protein S10	MRPS10	963.3	-0.87354	-2.01
224428_s_at	cell division cycle associated 7 /// cell division cycle associated 7	CDCA7	493.0	-1.42961	-2.01
204314_s_at	cAMP responsive element binding protein 1	CREB1	386.7	-1.66648	-2.01
203346_s_at	metal response element binding transcription factor 2	MTF2	340.6	-1.79665	-2.01
218701_at	lactamase, beta 2	LACTB2	748.0	-1.06570	-2.01
226528_at	metaxin 3	MTX3	468.9	-1.47716	-2.02
226272_at	Full length insert cDNA clone ZD79H10	---	561.8	-1.30952	-2.02
209054_s_at	Wolf-Hirschhorn syndrome candidate 1	WHSC1	480.2	-1.45434	-2.02
242648_at	kelch-like 8 ( <i>Drosophila</i> )	KLHL8	938.2	-0.89252	-2.02

230270_at	PRP38 pre-mRNA processing factor 38 (yeast) domain containing B	PRPF38B	698.6	-1.12142	-2.02
203903_s_at	hephaestin	HEPH	800.0	-1.01252	-2.03
227454_at	TAO kinase 1	TAOK1	951.6	-0.88231	-2.03
204018_x_at	hemoglobin, alpha 1 /// hemoglobin, alpha 1 /// hemoglobin, alpha 2 /// hemoglobin, alpha 2	HBA1 /// HBA2	415.9	-1.59369	-2.03
217777_s_at	butyrate-induced transcript 1	HSPC121	407.0	-1.61516	-2.03
1555058_a_at	lysophosphatidylglycerol acyltransferase 1	LPGAT1	445.1	-1.52720	-2.03
209358_at	TAF11 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 28kDa	TAF11	1057.6	-0.80822	-2.03
213007_at	hypothetical protein FLJ10719	FLJ10719	487.8	-1.43958	-2.03
209458_x_at	hemoglobin, alpha 1 /// hemoglobin, alpha 1 /// hemoglobin, alpha 2 /// hemoglobin, alpha 2	HBA1 /// HBA2	623.7	-1.21733	-2.03
240314_at	Nuclear receptor co-repressor 1	NCOR1	982.7	-0.85935	-2.04
230352_at	Phosphoribosyl pyrophosphate synthetase 2	PRPS2	620.3	-1.22213	-2.04
213899_at	methionyl aminopeptidase 2	METAP2	980.2	-0.86116	-2.04
206061_s_at	Dicer1, Dcr-1 homolog (Drosophila)	DICER1	1146.2	-0.75444	-2.04
211713_x_at	KIAA0101 /// KIAA0101	KIAA0101	724.3	-1.09179	-2.04
218622_at	nucleoporin 37kDa	NUP37	323.7	-1.85003	-2.04
222550_at	armadillo repeat containing 1	ARMC1	845.5	-0.96989	-2.04
221677_s_at	downstream neighbor of SON	DONSON	529.1	-1.36406	-2.04
218772_x_at	transmembrane protein 38B	TMEM38B	1203.4	-0.72297	-2.04
203358_s_at	enhancer of zeste homolog 2 (Drosophila)	EZH2	381.1	-1.68116	-2.05
213803_at	Karyopherin (importin) beta 1	KPNB1	899.9	-0.92309	-2.05
236985_at	Hypothetical protein PRO1843	EIF4B	847.4	-0.96820	-2.05
209484_s_at	chromosome 1 open reading frame 48	C1orf48	494.8	-1.42621	-2.05
214830_at	solute carrier family 38, member 6	SLC38A6	765.1	-1.04762	-2.06
205036_at	LSM6 homolog, U6 small nuclear RNA associated (S. cerevisiae)	LSM6	451.8	-1.51268	-2.06
201923_at	peroxiredoxin 4	PRDX4	804.5	-1.00810	-2.06
218993_at	RNA methyltransferase like 1	RNMTL1	357.1	-1.74751	-2.06
202626_s_at	v-yes-1 Yamaguchi sarcoma viral related oncogene homolog /// v-yes-1 Yamaguchi sarcoma viral related oncogene homolog	LYN	297.5	-1.93993	-2.06
227309_at	YOD1 OTU deubiquinating enzyme 1 homolog ( yeast)	YOD1	879.7	-0.93995	-2.06
218108_at	chromosome 14 open reading frame 130	C14orf130	962.8	-0.87395	-2.06
202020_s_at	LanC lantibiotic synthetase component C-like 1 (bacterial)	LANCL1	738.2	-1.07635	-2.06
201921_at	guanine nucleotide binding protein (G protein), gamma 10 /// hypothetical protein LOC552891	GNG10 /// LOC552891	777.6	-1.03484	-2.06

226803_at	chromatin modifying protein 4C	CHMP4C	372.2	-1.70518	-2.06
226151_x_at	crystallin, zeta (quinone reductase)-like 1	CRYZL1	349.2	-1.77072	-2.06
228652_at	hypothetical protein FLJ38288	FLJ38288	1008.3	-0.84123	-2.06
204444_at	kinesin family member 11	KIF11	726.5	-1.08936	-2.07
204779_s_at	homeo box B7	HOXB7	649.0	-1.18327	-2.07
205628_at	primase, polypeptide 2A, 58kDa	PRIM2A	1103.9	-0.77932	-2.07
215498_s_at	mitogen-activated protein kinase kinase 3 /// mitogen-activated protein kinase kinase 3	MAP2K3	802.2	-1.01037	-2.08
223191_at	chromosome 14 open reading frame 112	C14orf112	800.7	-1.01185	-2.08
213025_at	THUMP domain containing 1	THUMPD1	1289.1	-0.67996	-2.08
204023_at	replication factor C (activator 1) 4, 37kDa	RFC4	867.5	-0.95042	-2.08
201797_s_at	valyl-tRNA synthetase	VAR5	474.6	-1.46551	-2.08
209682_at	Cas-Br-M (murine) ecotropic retroviral transforming sequence b	CBLB	937.3	-0.89327	-2.08
209205_s_at	LIM domain only 4	LMO4	687.2	-1.13511	-2.09
208727_s_at	cell division cycle 42 (GTP binding protein, 25kDa)	CDC42	915.1	-0.91074	-2.09
238005_s_at	SIN3 homolog A, transcription regulator (yeast)	SIN3A	569.8	-1.29682	-2.09
210732_s_at	lectin, galactoside-binding, soluble, 8 (galectin 8)	LGALS8	644.2	-1.18957	-2.09
225925_s_at	ubiquitin specific peptidase 48	USP48	642.1	-1.19243	-2.09
1558956_s_at	WD repeat domain 56	WDR56	687.4	-1.13484	-2.09
230085_at	Transcribed locus	---	1014.3	-0.83710	-2.09
219439_at	core 1 synthase, glycoprotein-N-acetylglactosamine 3-beta-galactosyltransferase, 1	C1GALT1	762.0	-1.05087	-2.09
219581_at	tRNA splicing endonuclease 2 homolog (SEN2, S. cerevisiae)	TSEN2	575.5	-1.28793	-2.09
222889_at	DNA cross-link repair 1B (PSO2 homolog, S. cerevisiae)	DCLRE1B	943.7	-0.88830	-2.09
212686_at	protein phosphatase 1H (PP2C domain containing)	PPM1H	763.5	-1.04932	-2.09
229671_s_at	Chromosome 21 open reading frame 45	C21orf45	996.4	-0.84957	-2.10
202110_at	cytochrome c oxidase subunit VIIb	COX7B	1078.2	-0.79516	-2.10
1555878_at	Ribosomal protein S24	RPS24	934.8	-0.89517	-2.10
227798_at	SMAD, mothers against DPP homolog 1 (Drosophila)	SMAD1	900.5	-0.92262	-2.10
229097_at	Diaphanous homolog 3 (Drosophila)	DIAPH3	535.0	-1.35388	-2.11
222608_s_at	anillin, actin binding protein (scraps homolog, Drosophila)	ANLN	415.8	-1.59401	-2.11
223134_at	bobby sox homolog (Drosophila)	BBX	410.9	-1.60575	-2.11
203693_s_at	E2F transcription factor 3	E2F3	765.6	-1.04711	-2.11
1558166_at	hypothetical protein MGC16275	MGC16275	397.4	-1.63906	-2.11
210057_at	PI-3-kinase-related kinase SMG-1	SMG1	693.8	-1.12715	-2.11

218946_at	HIRA interacting protein 5	HIRIP5	1243.4	-0.70230	-2.11
213229_at	Dicer1, Dcr-1 homolog (Drosophila)	DICER1	478.7	-1.45736	-2.11
208846_s_at	voltage-dependent anion channel 3	VDAC3	708.7	-1.10960	-2.12
235984_at	---	---	845.6	-0.96981	-2.12
228400_at	Transcribed locus, moderately similar to NP_065910.1 Shroom-related protein; likely ortholog of mouse Shroom; F-actin-binding protein [Homo sapiens]	---	337.2	-1.80722	-2.12
213374_x_at	3-hydroxyisobutyryl-Coenzyme A hydrolase	HIBCH	979.6	-0.86160	-2.12
225674_at	B-cell receptor-associated protein 29	BCAP29	1129.6	-0.76404	-2.12
229744_at	Sperm specific antigen 2	SSFA2	803.6	-1.00898	-2.12
224512_s_at	hypothetical protein MGC14151 /// hypothetical protein MGC14151	MGC14151	507.1	-1.40326	-2.13
212798_s_at	ankyrin repeat and MYND domain containing 2	ANKMY2	810.6	-1.00224	-2.13
205194_at	phosphoserine phosphatase	PSPH	727.7	-1.08801	-2.13
202653_s_at	membrane-associated ring finger (C3HC4) 7	7-Mar	1162.1	-0.74545	-2.13
229704_at	Androgen-induced proliferation inhibitor	APRIN	377.7	-1.69037	-2.13
225073_at	periphilin 1	PPHLN1	951.6	-0.88229	-2.13
233587_s_at	signal-induced proliferation-associated 1 like 2	SIPA1L2	611.3	-1.23485	-2.13
201133_s_at	praja 2, RING-H2 motif containing	PJA2	774.7	-1.03778	-2.14
222423_at	Nedd4 family interacting protein 1	NDFIP1	719.2	-1.09759	-2.14
226694_at	PALM2-AKAP2 protein	PALM2-AKAP2	1164.3	-0.74421	-2.14
217414_x_at	hemoglobin, alpha 2	HBA2	1006.8	-0.84230	-2.14
235767_x_at	RNA U, small nuclear RNA export adaptor (phosphorylation regulated)	RNUXA	737.1	-1.07755	-2.14
224974_at	SDS3 protein	SDS3	1148.6	-0.75310	-2.14
1558922_at	TIA1 cytotoxic granule-associated RNA binding protein	TIA1	499.8	-1.41669	-2.14
221553_at	implantation-associated protein	DKFZp564K142	424.1	-1.57430	-2.14
208711_s_at	cyclin D1	CCND1	1045.3	-0.81628	-2.14
239346_at	Transcribed locus, weakly similar to XP_510104.1 PREDICTED: similar to hypothetical protein FLJ25224 [Pan troglodytes]	---	768.4	-1.04426	-2.15
222816_s_at	zinc finger, CCHC domain containing 2	ZCCHC2	907.9	-0.91652	-2.15
203016_s_at	synovial sarcoma, X breakpoint 2 interacting protein	SSX2IP	555.0	-1.32045	-2.15
213789_at	---	---	590.9	-1.26464	-2.15
223287_s_at	forkhead box P1	FOXP1	450.1	-1.51640	-2.15
222402_at	chromosome 13 open reading frame 12	C13orf12	882.7	-0.93744	-2.15

224827_at	Dendritic cell-derived ubiquitin-like protein	DC-UbP	1146.3	-0.75438	-2.15
222103_at	Activating transcription factor 1	ATF1	724.4	-1.09166	-2.15
223651_x_at	CDC23 (cell division cycle 23, yeast, homolog)	CDC23	1235.9	-0.70612	-2.16
222848_at	leucine zipper protein FKSG14	FKSG14	701.5	-1.11808	-2.16
223178_s_at	5'-nucleotidase, cytosolic II-like 1	NT5C2L1	1192.1	-0.72901	-2.16
223320_s_at	ATP-binding cassette, sub-family B (MDR/TAP), member 10	ABCB10	688.9	-1.13304	-2.16
223213_s_at	zinc fingers and homeoboxes 1	ZHX1	459.1	-1.49736	-2.16
226821_at	Full-length cDNA clone CS0DF029YD16 of Fetal brain of Homo sapiens (human)	---	937.8	-0.89289	-2.16
209065_at	ubiquinol-cytochrome c reductase binding protein	UQCRB	900.8	-0.92234	-2.16
219717_at	hypothetical protein FLJ20280	FLJ20280	421.6	-1.58027	-2.16
222850_s_at	DnaJ (Hsp40) homolog, subfamily B, member 14	DNAJB14	1216.1	-0.71632	-2.16
213883_s_at	TM2 domain containing 1	TM2D1	516.2	-1.38670	-2.17
223218_s_at	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, zeta	NFKBIZ	992.7	-0.85217	-2.17
223381_at	cell division cycle associated 1	CDCA1	965.8	-0.87168	-2.17
226851_at	lysophospholipase-like 1	LYPLAL1	641.1	-1.19373	-2.17
218549_s_at	family with sequence similarity 82, member B	FAM82B	968.4	-0.86977	-2.17
223239_at	chromosome 14 open reading frame 129	C14orf129	1264.4	-0.69187	-2.17
202468_s_at	catenin (cadherin-associated protein), alpha-like 1	CTNNAL1	630.5	-1.20806	-2.17
209865_at	solute carrier family 35 (UDP-N-acetylglucosamine (UDP-GlcNAc) transporter), member A3	SLC35A3	504.0	-1.40897	-2.18
213742_at	splicing factor, arginine/serine-rich 11	SFRS11	506.4	-1.40445	-2.18
217773_s_at	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 4, 9kDa	NDUFA4	388.1	-1.66275	-2.18
204146_at	RAD51 associated protein 1	RAD51AP1	947.6	-0.88533	-2.18
218982_s_at	mitochondrial ribosomal protein S17	MRPS17	993.1	-0.85187	-2.18
222714_s_at	lactamase, beta 2	LACTB2	480.2	-1.45444	-2.18
219097_x_at	hypothetical protein MGC2747	MGC2747	771.5	-1.04110	-2.19
225878_at	Kinesin family member 1B	KIF1B	403.7	-1.62332	-2.19
231913_s_at	chromosome X open reading frame 53	CXorf53	1600.1	-0.55534	-2.19
201437_s_at	eukaryotic translation initiation factor 4E	EIF4E	515.8	-1.38740	-2.19
229232_at	hypothetical protein FLJ36812	FLJ36812	1207.9	-0.72060	-2.19
204700_x_at	chromosome 1 open reading frame 107	C1orf107	741.1	-1.07319	-2.19
224450_s_at	RIO kinase 1 (yeast) /// RIO kinase 1 (yeast)	RIOK1	666.4	-1.16087	-2.20
204686_at	insulin receptor substrate 1	IRS1	387.1	-1.66545	-2.20
202451_at	general transcription factor IIIH,	GTF2H1	975.0	-0.86495	-2.20

	polypeptide 1, 62kDa				
210117_at	sperm associated antigen 1	SPAG1	750.6	-1.06295	-2.20
212989_at	transmembrane protein 23	TMEM23	1009.1	-0.84070	-2.20
233204_at	---	---	1035.2	-0.82296	-2.20
236678_at	Jagged 1 (Alagille syndrome)	JAG1	986.2	-0.85681	-2.20
229010_at	Cas-Br-M (murine) ecotropic retroviral transforming sequence	CBL	951.1	-0.88270	-2.20
225897_at	Myristoylated alanine-rich protein kinase C substrate	MARCKS	571.5	-1.29415	-2.21
218597_s_at	chromosome 10 open reading frame 70	C10orf70	844.7	-0.97062	-2.21
222243_s_at	transducer of ERBB2, 2	TOB2	1370.4	-0.64308	-2.21
221958_s_at	chromosome 1 open reading frame 139	C1orf139	624.1	-1.21685	-2.21
223292_s_at	mitochondrial ribosomal protein S15	MRPS15	581.8	-1.27823	-2.21
219148_at	PDZ binding kinase	PBK	883.4	-0.93680	-2.22
203880_at	COX17 homolog, cytochrome c oxidase assembly protein (yeast)	COX17	338.8	-1.80235	-2.22
209642_at	BUB1 budding uninhibited by benzimidazoles 1 homolog (yeast)	BUB1	408.7	-1.61096	-2.23
231809_x_at	programmed cell death 7	PDCD7	777.7	-1.03468	-2.23
232231_at	runt-related transcription factor 2	RUNX2	1202.1	-0.72366	-2.23
201867_s_at	transducin (beta)-like 1X-linked	TBL1X	1481.4	-0.59793	-2.23
227586_at	LOC124491	LOC124491	933.9	-0.89589	-2.24
218974_at	hypothetical protein FLJ10159	FLJ10159	717.3	-1.09971	-2.24
232338_at	Zinc finger protein 431	ZNF431	290.9	-1.96416	-2.25
220739_s_at	cyclin M3	CNNM3	759.3	-1.05372	-2.25
227107_at	Pannexin 1	PANX1	425.1	-1.57201	-2.26
238762_at	methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 2-like	MTHFD2L	1626.3	-0.54659	-2.26
213647_at	DNA2 DNA replication helicase 2-like (yeast)	DNA2L	677.0	-1.14755	-2.26
211699_x_at	hemoglobin, alpha 1 /// hemoglobin, alpha 1 /// hemoglobin, alpha 2 /// hemoglobin, alpha 2	HBA1 /// HBA2	1103.7	-0.77941	-2.27
1555247_a_at	Rap guanine nucleotide exchange factor (GEF) 6	RAPGEF6	579.0	-1.28265	-2.27
218350_s_at	geminin, DNA replication inhibitor	GMNN	467.1	-1.48080	-2.27
205264_at	CD3E antigen, epsilon polypeptide associated protein	CD3EAP	1100.1	-0.78162	-2.27
213705_at	Methionine adenosyltransferase II, alpha	MAT2A	498.8	-1.41859	-2.27
222014_x_at	mitochondrial translation optimization 1 homolog (S. cerevisiae)	MTO1	950.1	-0.88348	-2.27
222396_at	hematological and neurological expressed 1	HN1	725.8	-1.09010	-2.27
227085_at	H2A histone family, member V	H2AFV	898.0	-0.92463	-2.27
228401_at	ATPase family, AAA domain containing 2	ATAD2	789.2	-1.02314	-2.27
209451_at	TRAF family member-associated NFKB	TANK	400.4	-1.63145	-2.28

	activator					
53968_at	KIAA1698 protein	KIAA1698	591.6	-1.26346	-2.28	
226837_at	sprouty-related, EVH1 domain containing 1	SPRED1	575.4	-1.28820	-2.28	
209535_s_at	---	---	699.2	-1.12076	-2.28	
207785_s_at	recombining binding protein suppressor of hairless (Drosophila)	RBPSUH	963.8	-0.87314	-2.28	
227632_at	KIAA1171 protein	KIAA1171	1241.5	-0.70330	-2.28	
230263_s_at	dedicator of cytokinesis 5	DOCK5	1523.3	-0.58230	-2.28	
232198_at	CDNA FLJ12676 fis, clone NT2RM4002383	---	860.8	-0.95627	-2.28	
226981_at	Similar to CDNA sequence BC021608	LOC143941	1223.7	-0.71236	-2.29	
201129_at	splicing factor, arginine/serine-rich 7, 35kDa	SFRS7	1026.8	-0.82856	-2.29	
201877_s_at	protein phosphatase 2, regulatory subunit B (B56), gamma isoform	PPP2R5C	671.3	-1.15472	-2.29	
226921_at	Ubiquitin protein ligase E3 component n-recognin 1	UBR1	915.6	-0.91032	-2.29	
203544_s_at	signal transducing adaptor molecule (SH3 domain and ITAM motif) 1	STAM	345.2	-1.78283	-2.29	
204867_at	GTP cyclohydrolase I feedback regulator	GCHFR	620.3	-1.22205	-2.29	
227682_at	Transcribed locus	---	566.1	-1.30277	-2.29	
209568_s_at	ral guanine nucleotide dissociation stimulator-like 1	RGL1	859.0	-0.95783	-2.29	
227105_at	centrosome and spindle pole associated protein 1	CSPP1	407.8	-1.61310	-2.29	
215726_s_at	cytochrome b-5	CYB5	806.3	-1.00640	-2.30	
202184_s_at	nucleoporin 133kDa	NUP133	1133.9	-0.76153	-2.30	
212840_at	KIAA0794 protein	KIAA0794	1247.6	-0.70021	-2.30	
228220_at	FCH domain only 2	FCHO2	1540.0	-0.57623	-2.30	
227786_at	thyroid hormone receptor associated protein 6	THRAP6	635.2	-1.20162	-2.30	
1552426_a_at	TM2 domain containing 3	TM2D3	806.1	-1.00663	-2.30	
204510_at	CDC7 cell division cycle 7 (S. cerevisiae)	CDC7	684.4	-1.13855	-2.31	
203213_at	Cell division cycle 2, G1 to S and G2 to M	CDC2	1279.4	-0.68457	-2.31	
200783_s_at	stathmin 1/oncoprotein 18	STMN1	1381.9	-0.63810	-2.32	
226860_at	transmembrane protein 19	TMEM19	821.0	-0.99238	-2.32	
234986_at	Glutamate-cysteine ligase, modifier subunit	GCLM	407.3	-1.61432	-2.32	
225686_at	family with sequence similarity 33, member A	FAM33A	1388.4	-0.63536	-2.32	
220864_s_at	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 13	NDUFA13	1090.1	-0.78774	-2.32	
212573_at	KIAA0830 protein	KIAA0830	775.8	-1.03664	-2.32	
226931_at	ARG99 protein	ARG99	789.0	-1.02336	-2.32	
218459_at	torsin family 3, member A	TOR3A	981.8	-0.86001	-2.32	
203359_s_at	c-myc binding protein	MYCBP	1257.5	-0.69527	-2.33	

228988_at	zinc finger protein 6 (CMPX1)	ZNF6	1775.3	-0.50099	-2.33
212604_at	mitochondrial ribosomal protein S31	MRPS31	1075.7	-0.79673	-2.33
222607_s_at	KIAA1008	KIAA1008	1235.3	-0.70642	-2.33
219512_at	chromosome 20 open reading frame 172	C20orf172	1324.3	-0.66355	-2.33
226936_at	chromosome 6 open reading frame 173	C6orf173	1046.0	-0.81582	-2.33
206102_at	DNA replication complex GINS protein PSF1	PSF1	566.5	-1.30212	-2.34
203176_s_at	transcription factor A, mitochondrial	TFAM	1444.5	-0.61234	-2.34
226364_at	Huntingtin interacting protein 1	HIP1	875.9	-0.94319	-2.34
202353_s_at	proteasome (prosome, macropain) 26S subunit, non-ATPase, 12	PSMD12	763.4	-1.04946	-2.34
226283_at	WD repeat domain 51B	WDR51B	1101.6	-0.78068	-2.34
202904_s_at	LSM5 homolog, U6 small nuclear RNA associated ( <i>S. cerevisiae</i> )	LSM5	571.7	-1.29382	-2.34
1557987_at	PI-3-kinase-related kinase SMG-1 - like locus	LOC641298	350.1	-1.76812	-2.34
222533_at	cereblon	CRBN	1584.4	-0.56068	-2.35
224837_at	forkhead box P1	FOXP1	621.4	-1.22063	-2.35
228280_at	similar to RIKEN cDNA 1200014N16 gene	MGC14289	1386.9	-0.63597	-2.35
218014_at	pericentrin 1	PCNT1	517.2	-1.38490	-2.35
209932_s_at	dUTP pyrophosphatase	DUT	1306.3	-0.67183	-2.35
52285_f_at	chromosome 18 open reading frame 9	C18orf9	847.1	-0.96843	-2.35
218602_s_at	family with sequence similarity 29, member A	FAM29A	396.8	-1.64056	-2.36
223281_s_at	COX15 homolog, cytochrome c oxidase assembly protein (yeast)	COX15	1571.2	-0.56522	-2.36
1552310_at	hypothetical protein MGC29937	MGC29937	466.6	-1.48185	-2.36
238002_at	golgi phosphoprotein 4	GOLPH4	965.2	-0.87215	-2.37
1569433_at	SAM domain containing 1	LOC389432	764.8	-1.04801	-2.37
209512_at	hydroxysteroid dehydrogenase like 2	HSDL2	396.5	-1.64134	-2.37
204252_at	cyclin-dependent kinase 2	CDK2	778.5	-1.03393	-2.37
224802_at	Nedd4 family interacting protein 2	NDFIP2	1144.4	-0.75549	-2.37
227110_at	heterogeneous nuclear ribonucleoprotein C (C1/C2)	HNRPC	330.5	-1.82813	-2.38
205596_s_at	SMAD specific E3 ubiquitin protein ligase 2	SMURF2	793.6	-1.01878	-2.38
219258_at	timeless-interacting protein	FLJ20516	720.3	-1.09626	-2.38
213344_s_at	H2A histone family, member X	H2AFX	469.8	-1.47519	-2.38
203560_at	gamma-glutamyl hydrolase (conjugase, foylypolygammaglutamyl hydrolase)	GGH	1108.1	-0.77675	-2.39
206272_at	S-phase response (cyclin-related)	SPHAR	1899.0	-0.46764	-2.39
225688_s_at	pleckstrin homology-like domain, family B, member 2	PHLDB2	581.5	-1.27878	-2.39
203124_s_at	solute carrier family 11 (proton-coupled divalent metal ion transporters), member 2	SLC11A2	728.8	-1.08680	-2.39

203987_at	frizzled homolog 6 ( <i>Drosophila</i> )	FZD6	586.0	-1.27199	-2.39
212106_at	expressed in T-cells and eosinophils in atopic dermatitis	ETEA	1628.3	-0.54593	-2.40
218311_at	mitogen-activated protein kinase kinase kinase 3	MAP4K3	878.6	-0.94091	-2.40
203401_at	phosphoribosyl pyrophosphate synthetase 2	PRPS2	886.5	-0.93418	-2.40
208719_s_at	DEAD (Asp-Glu-Ala-Asp) box polypeptide 17	DDX17	868.3	-0.94973	-2.40
235296_at	eukaryotic translation initiation factor 5A2	EIF5A2	745.7	-1.06822	-2.40
212398_at	radixin	RDX	1353.0	-0.65065	-2.40
1560296_at	Dystonin	DST	602.4	-1.24757	-2.40
225633_at	dpy-19-like 3 ( <i>C. elegans</i> )	DPY19L3	613.5	-1.23166	-2.41
202660_at	Family with sequence similarity 20, member C	ITPR2	1165.2	-0.74375	-2.41
242787_at	---	---	1492.8	-0.59362	-2.41
222843_at	fidgetin-like 1	FIGNL1	884.9	-0.93554	-2.41
91816_f_at	ring finger and KH domain containing 1	RKHD1	484.0	-1.44701	-2.41
227164_at	Splicing factor, arginine/serine-rich 1 (splicing factor 2, alternate splicing factor)	SFRS1	787.4	-1.02490	-2.42
227740_at	U2AF homology motif (UHM) kinase 1	UHMK1	1786.5	-0.49783	-2.42
213246_at	chromosome 14 open reading frame 109	C14orf109	1033.1	-0.82433	-2.42
1560116_a_at	neural precursor cell expressed, developmentally down-regulated 1	NEDD1	1877.3	-0.47324	-2.43
226022_at	SAM and SH3 domain containing 1	SASH1	919.6	-0.90718	-2.43
209224_s_at	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 2, 8kDa	NDUFA2	1537.2	-0.57724	-2.43
212387_at	Transcription factor 4	TCF4	634.9	-1.20199	-2.43
206928_at	zinc finger protein 124 (HZF-16)	ZNF124	472.6	-1.46955	-2.43
221046_s_at	HSPC135 protein	HSPC135	1135.7	-0.76051	-2.43
236696_at	U2-associated SR140 protein	SR140	1322.5	-0.66437	-2.44
204918_s_at	myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, <i>Drosophila</i> ); translocated to, 3	MLLT3	2046.8	-0.43225	-2.44
219335_at	armadillo repeat containing, X-linked 5	ARMCX5	1689.7	-0.52639	-2.45
201316_at	proteasome (prosome, macropain) subunit, alpha type, 2	PSMA2	1518.0	-0.58424	-2.45
212511_at	phosphatidylinositol binding clathrin assembly protein	PICALM	626.2	-1.21398	-2.46
229193_at	Cisplatin resistance-associated overexpressed protein	CROP	312.3	-1.88800	-2.46
204839_at	processing of precursor 5, ribonuclease P/MRP subunit ( <i>S. cerevisiae</i> )	POP5	691.6	-1.12980	-2.46
203803_at	prenylcysteine oxidase 1	PCYOX1	969.8	-0.86873	-2.46
229442_at	chromosome 18 open reading frame 54	C18orf54	1223.1	-0.71265	-2.46
225583_at	UDP-glucuronate decarboxylase 1	UXS1	1649.3	-0.53911	-2.47
203225_s_at	riboflavin kinase	RFK	735.6	-1.07918	-2.47

225223_at	SMAD, mothers against DPP homolog 5 (Drosophila)	SMAD5	1901.7	-0.46697	-2.47
218875_s_at	F-box protein 5	FBXO5	937.8	-0.89287	-2.47
201448_at	TIA1 cytotoxic granule-associated RNA binding protein	TIA1	282.8	-1.99479	-2.47
217989_at	dehydrogenase/reductase (SDR family) member 8	DHRS8	874.3	-0.94455	-2.48
235099_at	chemokine-like factor superfamily 8	CKLFSF8	1409.2	-0.62666	-2.48
1558080_s_at	hypothetical protein LOC144871	LOC144871	1417.1	-0.62339	-2.48
235474_at	Transcribed locus, weakly similar to XP_331555.1 hypothetical protein [Neurospora crassa]	---	924.3	-0.90341	-2.48
1558048_x_at	---	---	1222.4	-0.71304	-2.48
219016_at	hypothetical protein FLJ13149	FLJ13149	748.5	-1.06517	-2.48
225836_s_at	hypothetical protein MGC13204	MGC13204	308.3	-1.90182	-2.48
219161_s_at	chemokine-like factor	CKLF	600.4	-1.25056	-2.49
201670_s_at	myristoylated alanine-rich protein kinase C substrate	MARCKS	1029.0	-0.82710	-2.49
229309_at	Adrenergic, beta-1-, receptor	ADRB1	663.3	-1.16479	-2.49
201110_s_at	thrombospondin 1	THBS1	789.9	-1.02246	-2.50
213226_at	Cyclin A2	CCNA2	564.3	-1.30558	-2.50
1555920_at	Chromobox homolog 3 (HP1 gamma homolog, Drosophila)	CBX3	1542.3	-0.57543	-2.50
232103_at	3'(2'), 5'-bisphosphate nucleotidase 1	BPNT1	1299.4	-0.67509	-2.50
228106_at	hypothetical protein FLJ20280	FLJ20280	1406.6	-0.62773	-2.50
226479_at	kelch repeat and BTB (POZ) domain containing 6	KBTBD6	1008.7	-0.84096	-2.50
226780_s_at	hypothetical protein HSPC268	HSPC268	1404.2	-0.62872	-2.51
1558142_at	trinucleotide repeat containing 6B	TNRC6B	1138.5	-0.75887	-2.51
226302_at	Transcribed locus	---	1405.7	-0.62808	-2.51
238852_at	Paired related homeobox 1	PRRX1	1184.3	-0.73321	-2.52
202960_s_at	methylmalonyl Coenzyme A mutase	MUT	840.7	-0.97424	-2.52
225536_at	transmembrane protein 54	TMEM54	1067.6	-0.80180	-2.52
217526_at	nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 2 interacting protein	NFATC2IP	589.3	-1.26690	-2.52
208956_x_at	dUTP pyrophosphatase	DUT	1487.7	-0.59555	-2.52
226287_at	NY-REN-41 antigen	NY-REN-41	686.3	-1.13614	-2.53
228446_at	KIAA2026	KIAA2026	1924.7	-0.46116	-2.53
209434_s_at	phosphoribosyl pyrophosphate amidotransferase	PPAT	792.5	-1.01989	-2.53
236356_at	NADH dehydrogenase (ubiquinone) Fe-S protein 1, 75kDa (NADH-coenzyme Q reductase)	NDUFS1	850.7	-0.96517	-2.53
224861_at	Guanine nucleotide binding protein (G protein), q polypeptide	GNAQ	401.7	-1.62808	-2.53
239046_at	Transcribed locus	---	1445.7	-0.61186	-2.54

218254_s_at	SAR1 gene homolog B ( <i>S. cerevisiae</i> )	SAR1B	906.9	-0.91738	-2.55
218458_at	germ cell-less homolog 1 ( <i>Drosophila</i> )	GCL	783.4	-1.02897	-2.55
212168_at	RNA binding motif protein 12	RBM12	1210.9	-0.71901	-2.56
205443_at	small nuclear RNA activating complex, polypeptide 1, 43kDa	SNAPC1	473.0	-1.46878	-2.56
227936_at	transmembrane protein 68	TMEM68	1231.3	-0.70848	-2.56
221543_s_at	SPFH domain family, member 2	SPFH2	2312.9	-0.37823	-2.56
210222_s_at	reticulon 1	RTN1	1344.0	-0.65464	-2.58
235088_at	hypothetical protein LOC201725	LOC201725	848.6	-0.96705	-2.59
201143_s_at	eukaryotic translation initiation factor 2, subunit 1 alpha, 35kDa	EIF2S1	1351.4	-0.65135	-2.59
222039_at	hypothetical protein LOC146909	LOC146909	318.2	-1.86826	-2.60
227932_at	ariadne homolog 2 ( <i>Drosophila</i> )	ARIH2	1639.2	-0.54237	-2.60
242617_at	Transmembrane emp24 protein transport domain containing 8	TMED8	1386.2	-0.63627	-2.61
229787_s_at	O-linked N-acetylglucosamine (GlcNAc) transferase (UDP-N-acetylglucosamine:polypeptide-N-acetylglucosaminyl transferase)	OGT	1528.4	-0.58041	-2.61
225412_at	Hypothetical protein FLJ14681	FLJ14681	979.5	-0.86166	-2.61
227931_at	MRNA; cDNA DKFZp686D22106 (from clone DKFZp686D22106)	---	1800.8	-0.49382	-2.61
227350_at	CDNA FLJ11381 fis, clone HEMBA1000501	---	1367.1	-0.64448	-2.62
201363_s_at	influenza virus NS1A binding protein	IVNS1ABP	867.3	-0.95060	-2.62
218947_s_at	PAP associated domain containing 1	PAPD1	558.3	-1.31520	-2.62
225835_at	solute carrier family 12 (sodium/potassium/chloride transporters), member 2	SLC12A2	985.7	-0.85718	-2.63
214843_s_at	ubiquitin specific peptidase 33	USP33	1681.0	-0.52909	-2.63
238058_at	hypothetical protein LOC150381	LOC150381	1123.4	-0.76765	-2.63
235177_at	similar to hepatocellular carcinoma-associated antigen HCA557b	LOC151194	924.7	-0.90313	-2.64
239355_at	Germ cell-less homolog 1 ( <i>Drosophila</i> )	GCL	1097.4	-0.78328	-2.64
232238_at	asp (abnormal spindle)-like, microcephaly associated ( <i>Drosophila</i> )	ASPM	1680.7	-0.52919	-2.64
212918_at	RecQ protein-like (DNA helicase Q1-like)	RECQL	794.8	-1.01765	-2.65
240636_at	Hypothetical protein FLJ23342	FLJ23342	359.0	-1.74211	-2.65
227630_at	Protein phosphatase 2, regulatory subunit B (B56), epsilon isoform	PPP2R5E	886.1	-0.93456	-2.65
211075_s_at	CD47 antigen (Rh-related antigen, integrin-associated signal transducer) /// CD47 antigen (Rh-related antigen, integrin-associated signal transducer)	CD47	728.3	-1.08732	-2.65
213002_at	Myristoylated alanine-rich protein kinase C substrate	MARCKS	1702.7	-0.52240	-2.65
213410_at	chromosome 10 open reading frame 137	C10orf137	1311.5	-0.66941	-2.66

209422_at	PHD finger protein 20	PHF20	1119.8	-0.76977	-2.66
222679_s_at	DCN1, defective in cullin neddylation 1, domain containing 1 ( <i>S. cerevisiae</i> )	DCUN1D1	1022.1	-0.83173	-2.67
235390_at	p18 splicing regulatory protein	P18SRP	792.0	-1.02034	-2.67
219031_s_at	comparative gene identification transcript 37	CGI-37	648.8	-1.18348	-2.67
225655_at	ubiquitin-like, containing PHD and RING finger domains, 1	UHRF1	786.8	-1.02559	-2.67
221548_s_at	integrin-linked kinase-associated serine/threonine phosphatase 2C	ILKAP	549.2	-1.33000	-2.68
218395_at	ARP6 actin-related protein 6 homolog (yeast)	ACTR6	1181.0	-0.73501	-2.68
201513_at	translin	TSN	1638.6	-0.54256	-2.68
209780_at	putative homeodomain transcription factor 2	PHTF2	1410.3	-0.62618	-2.69
212057_at	KIAA0182 protein	KIAA0182	1820.7	-0.48833	-2.69
212558_at	sprouty homolog 1, antagonist of FGF signaling ( <i>Drosophila</i> )	SPRY1	2019.1	-0.43856	-2.69
34764_at	leucyl-tRNA synthetase 2, mitochondrial	LARS2	1819.3	-0.48871	-2.71
214157_at	GNAS complex locus	GNAS	427.3	-1.56695	-2.71
206860_s_at	hypothetical protein FLJ20323	FLJ20323	660.6	-1.16816	-2.71
223515_s_at	coenzyme Q3 homolog, methyltransferase (yeast)	COQ3	1338.0	-0.65734	-2.72
218734_at	hypothetical protein FLJ13848	FLJ13848	1624.9	-0.54706	-2.72
212944_at	Mitochondrial ribosomal protein S6	MRPS6	727.1	-1.08862	-2.73
228221_at	solute carrier family 44, member 3	SLC44A3	762.7	-1.05022	-2.73
234947_s_at	chromosome 10 open reading frame 84	C10orf84	2685.6	-0.31820	-2.73
224838_at	forkhead box P1	FOXP1	1784.2	-0.49845	-2.73
204900_x_at	sin3-associated polypeptide, 30kDa	SAP30	833.5	-0.98077	-2.73
203972_s_at	peroxisomal biogenesis factor 3	PEX3	1361.9	-0.64674	-2.73
226753_at	family with sequence similarity 76, member B	FAM76B	2026.8	-0.43678	-2.74
225343_at	transmembrane emp24 protein transport domain containing 8	TMED8	2036.2	-0.43464	-2.74
219497_s_at	B-cell CLL/lymphoma 11A (zinc finger protein)	BCL11A	1371.7	-0.64252	-2.74
201668_x_at	myristoylated alanine-rich protein kinase C substrate	MARCKS	1662.9	-0.53479	-2.75
218117_at	ring-box 1	RBX1	1433.0	-0.61693	-2.75
238756_at	Growth arrest-specific 2 like 3	GAS2L3	651.6	-1.17979	-2.75
214124_x_at	FGFR1 oncogene partner	FGFR1OP	1661.6	-0.53519	-2.76
228654_at	hypothetical protein LOC139886	LOC139886	1941.1	-0.45710	-2.76
1554768_a_at	MAD2 mitotic arrest deficient-like 1 (yeast)	MAD2L1	491.2	-1.43310	-2.77
218585_s_at	denticleless homolog ( <i>Drosophila</i> )	DTL	1166.1	-0.74322	-2.77
213571_s_at	eukaryotic translation initiation factor 4E member 2	EIF4E2	2127.8	-0.41459	-2.77

237400_at	ATP synthase, H <sup>+</sup> transporting, mitochondrial F0 complex, subunit s (factor B)	ATP5S	1318.2	-0.66633	-2.78
235341_at	Hypothetical protein LOC144871	DNAJC3	1146.8	-0.75408	-2.78
219363_s_at	MTERF domain containing 1	MTERFD1	2316.1	-0.37765	-2.78
201602_s_at	protein phosphatase 1, regulatory (inhibitor) subunit 12A	PPP1R12A	2024.2	-0.43737	-2.79
53071_s_at	hypothetical protein FLJ22222	FLJ22222	1905.3	-0.46604	-2.79
1557915_s_at	glutathione S-transferase omega 1	GSTO1	1743.7	-0.51013	-2.79
231832_at	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 4 (GalNAc-T4)	GALNT4	2155.3	-0.40886	-2.79
229966_at	Ewing sarcoma breakpoint region 1	EWSR1	1235.6	-0.70626	-2.79
205176_s_at	integrin beta 3 binding protein (beta3-endonexin)	ITGB3BP	844.2	-0.97106	-2.79
225799_at	hypothetical protein MGC4677 /// hypothetical LOC541471	MGC4677 /// LOC541471	701.8	-1.11763	-2.80
203625_x_at	S-phase kinase-associated protein 2 (p45)	SKP2	924.9	-0.90292	-2.82
227653_at	TRM5 tRNA methyltransferase 5 homolog (S. cerevisiae)	TRMT5	1029.8	-0.82659	-2.82
203989_x_at	coagulation factor II (thrombin) receptor	F2R	686.3	-1.13623	-2.83
201280_s_at	disabled homolog 2, mitogen-responsive phosphoprotein (Drosophila)	DAB2	1320.3	-0.66537	-2.83
232021_at	hypothetical protein LOC283464	LOC283464	1781.3	-0.49928	-2.84
230009_at	Hypothetical protein FLJ21103	FLJ21103	774.5	-1.03794	-2.85
206052_s_at	stem-loop (histone) binding protein	SLBP	459.5	-1.49653	-2.85
239082_at	CDNA clone IMAGE:5311370	---	1554.5	-0.57107	-2.85
225730_s_at	THUMP domain containing 3	THUMPD3	1418.6	-0.62279	-2.86
227223_at	RNA-binding region (RNP1, RRM) containing 2	RNPC2	526.9	-1.36782	-2.87
201896_s_at	proline/serine-rich coiled-coil 1	PSRC1	1471.4	-0.60177	-2.88
213373_s_at	caspase 8, apoptosis-related cysteine peptidase	CASP8	1539.0	-0.57660	-2.88
212142_at	MCM4 minichromosome maintenance deficient 4 (S. cerevisiae)	MCM4	1259.2	-0.69442	-2.88
204798_at	v-myb myeloblastosis viral oncogene homolog (avian)	MYB	578.4	-1.28353	-2.88
228731_at	CDNA: FLJ21462 fis, clone COL04744	---	2992.1	-0.27867	-2.88
209681_at	solute carrier family 19 (thiamine transporter), member 2	SLC19A2	1732.8	-0.51335	-2.88
207717_s_at	plakophilin 2	PKP2	805.7	-1.00700	-2.88
1553106_at	hypothetical protein FLJ37562	FLJ37562	2157.9	-0.40832	-2.88
218392_x_at	sideroflexin 1	SFXN1	428.8	-1.56352	-2.88
227787_s_at	thyroid hormone receptor associated protein 6	THRAP6	349.6	-1.76943	-2.88

203566_s_at	amylo-1, 6-glucosidase, 4-alpha-glucanotransferase (glycogen debranching enzyme, glycogen storage disease type III)	AGL	1744.7	-0.50983	-2.90
214383_x_at	kelch domain containing 3	KLHDC3	629.8	-1.20902	-2.91
209210_s_at	pleckstrin homology domain containing, family C (with FERM domain) member 1	PLEKHC1	1475.0	-0.60038	-2.91
213701_at	hypothetical protein DKFZp434N2030	DKFZp434N2030	1402.8	-0.62932	-2.94
222791_at	round spermatid basic protein 1	RSBN1	1378.0	-0.63978	-2.94
231174_s_at	Erythrocyte membrane protein band 4.1-like 2	EPB41L2	870.6	-0.94772	-2.95
231862_at	Chromobox homolog 5 (HP1 alpha homolog, Drosophila)	CBX5	2038.7	-0.43407	-2.95
201845_s_at	RING1 and YY1 binding protein	RYBP	705.7	-1.11314	-2.96
235003_at	U2AF homology motif (UHM) kinase 1	UHMK1	1765.4	-0.50381	-2.97
219493_at	SHC SH2-domain binding protein 1	SHCBP1	1720.1	-0.51714	-2.99
227379_at	O-acyltransferase (membrane bound) domain containing 1	OACT1	1524.9	-0.58171	-2.99
225181_at	AT rich interactive domain 1B (SWI1-like)	ARID1B	1786.1	-0.49792	-2.99
214239_x_at	polycomb group ring finger 2	PCGF2	2136.3	-0.41281	-3.00
202620_s_at	procollagen-lysine, 2-oxoglutarate 5-dioxygenase 2	PLOD2	1638.1	-0.54274	-3.01
239022_at	Succinate dehydrogenase complex, subunit A, flavoprotein-like 2	SDHAL2	1275.7	-0.68635	-3.02
213251_at	Hypothetical LOC 441046	---	444.9	-1.52757	-3.02
212544_at	zinc finger, HIT type 3	ZNHIT3	917.5	-0.90880	-3.02
225922_at	Hypothetical protein FLJ25371	FLJ25371	1980.9	-0.44748	-3.03
218397_at	Fanconi anemia, complementation group L	FANCL	378.9	-1.68715	-3.04
220085_at	helicase, lymphoid-specific	HELLS	876.3	-0.94285	-3.04
221558_s_at	lymphoid enhancer-binding factor 1	LEF1	402.8	-1.62557	-3.04
213637_at	Transcribed locus, moderately similar to XP_517655.1 PREDICTED: similar to KIAA0825 protein [Pan troglodytes]	---	1792.1	-0.49625	-3.05
203789_s_at	sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3C	SEMA3C	2444.4	-0.35526	-3.05
230060_at	cell division cycle associated 7	CDCA7	2146.2	-0.41074	-3.07
209527_at	exosome component 2	EXOSC2	1807.7	-0.49191	-3.09
1553677_a_at	TIP41, TOR signalling pathway regulator-like ( <i>S. cerevisiae</i> )	TIPRL	1041.4	-0.81881	-3.09
201339_s_at	sterol carrier protein 2	SCP2	812.0	-1.00097	-3.10
225060_at	low density lipoprotein receptor-related protein 11	LRP11	587.6	-1.26952	-3.10
209377_s_at	high mobility group nucleosomal binding domain 3	HMGN3	1721.2	-0.51679	-3.10
212361_s_at	ATPase, Ca <sup>++</sup> transporting, cardiac muscle, slow twitch 2	ATP2A2	2724.8	-0.31271	-3.11

212240_s_at	phosphoinositide-3-kinase, regulatory subunit 1 (p85 alpha)	PIK3R1	1796.2	-0.49509	-3.12
213694_at	round spermatid basic protein 1	RSBN1	1699.9	-0.52324	-3.12
214052_x_at	BAT2 domain containing 1	XTP2	1006.3	-0.84264	-3.13
242214_at	ribosomal protein S27a /// similar to bA92K2.2 (similar to ubiquitin) /// similar to ribosomal protein S27a	RPS27A /// LOC388720 /// LOC389425	828.5	-0.98544	-3.13
204033_at	thyroid hormone receptor interactor 13	TRIP13	676.4	-1.14837	-3.14
225040_s_at	ribulose-5-phosphate-3-epimerase	RPE	1177.1	-0.73715	-3.18
218577_at	leucine rich repeat containing 40	LRRC40	2961.3	-0.28232	-3.18
203628_at	insulin-like growth factor 1 receptor	IGF1R	2131.7	-0.41379	-3.18
212861_at	hypothetical protein MGC11308	MGC11308	460.8	-1.49383	-3.18
205909_at	polymerase (DNA directed), epsilon 2 (p59 subunit)	POLE2	986.5	-0.85664	-3.18
222787_s_at	hypothetical protein FLJ11273	FLJ11273	2869.8	-0.29356	-3.18
238431_at	Transcribed locus, weakly similar to NP_055301.1 neuronal thread protein AD7c-NTP [Homo sapiens]	---	1729.7	-0.51428	-3.18
204835_at	polymerase (DNA directed), alpha	POLA	896.4	-0.92596	-3.18
228694_at	Homo sapiens, clone IMAGE:3352913, mRNA	---	2228.3	-0.39422	-3.19
223434_at	guanylate binding protein 3	GBP3	1658.1	-0.53631	-3.20
212215_at	prolyl endopeptidase-like	PREPL	1730.9	-0.51391	-3.20
205063_at	survival of motor neuron protein interacting protein 1	SIP1	1629.8	-0.54546	-3.21
216450_x_at	tumor rejection antigen (gp96) 1	TRA1	971.0	-0.86785	-3.21
222487_s_at	ribosomal protein S27-like	RPS27L	1589.9	-0.55877	-3.22
225455_at	transcriptional adaptor 1 (HFI1 homolog, yeast)-like	TADA1L	3249.6	-0.25056	-3.22
213891_s_at	Transcription factor 4	TCF4	1356.9	-0.64893	-3.23
212731_at	ankyrin repeat domain 46	ANKRD46	2327.3	-0.37560	-3.23
227503_at	CDNA FLJ43100 fis, clone CTONG2003100	---	634.6	-1.20245	-3.23
1567213_at	pinin, desmosome associated protein	PNN	1238.3	-0.70489	-3.23
226420_at	ecotropic viral integration site 1	EVI1	981.5	-0.86024	-3.24
201968_s_at	phosphoglucomutase 1	PGM1	2252.3	-0.38958	-3.24
223210_at	churchill domain containing 1	CHURC1	696.8	-1.12355	-3.25
222803_at	phosphoribosyl transferase domain containing 1	PRTFDC1	834.4	-0.97995	-3.25
218236_s_at	protein kinase D3	PRKD3	1289.6	-0.67968	-3.28
218166_s_at	hepatitis B virus x associated protein	HBXAP	1140.8	-0.75751	-3.30
212543_at	absent in melanoma 1	AIM1	2813.8	-0.30075	-3.31
226085_at	Chromobox homolog 5 (HP1 alpha homolog, Drosophila)	CBX5	2309.8	-0.37880	-3.31
209754_s_at	thymopoietin	TMPO	2263.0	-0.38754	-3.32

222600_s_at	hypothetical protein FLJ10808	FLJ10808	1478.0	-0.59925	-3.33
213032_at	Nuclear factor I/B	NFIB	2105.0	-0.41945	-3.33
226980_at	DEP domain containing 1B	DEPDC1B	2734.3	-0.31141	-3.36
219498_s_at	B-cell CLL/lymphoma 11A (zinc finger protein)	BCL11A	3191.8	-0.25653	-3.37
235484_at	protein prenyltransferase alpha subunit repeat containing 1	PTAR1	2834.7	-0.29804	-3.38
224352_s_at	cofilin 2 (muscle) /// cofilin 2 (muscle)	CFL2	857.3	-0.95933	-3.38
213094_at	G protein-coupled receptor 126	GPR126	3302.6	-0.24525	-3.38
205296_at	retinoblastoma-like 1 (p107)	RBL1	568.8	-1.29840	-3.40
224755_at	SM-11044 binding protein	SMBP	397.2	-1.63934	-3.43
225336_at	---	---	1202.1	-0.72369	-3.45
228915_at	dachshund homolog 1 (Drosophila)	DACH1	1958.8	-0.45279	-3.46
219004_s_at	chromosome 21 open reading frame 45	C21orf45	533.0	-1.35724	-3.48
230560_at	syntaxin binding protein 6 (amisyn)	STXBP6	3015.7	-0.27592	-3.49
223059_s_at	chromosome 10 open reading frame 45	C10orf45	2496.7	-0.34670	-3.50
228357_at	zinc finger CCCH-type containing 5	ZC3H5	1611.4	-0.55152	-3.50
224595_at	solute carrier family 44, member 1	SLC44A1	975.2	-0.86480	-3.50
205943_at	tryptophan 2,3-dioxygenase	TDO2	4577.9	-0.15146	-3.51
227038_at	hypothetical protein MGC26963	MGC26963	1950.9	-0.45471	-3.54
211406_at	immediate early response 3 interacting protein 1	IER3IP1	3556.0	-0.22184	-3.55
242890_at	Helicase, lymphoid-specific	HELLS	1522.0	-0.58277	-3.56
204299_at	FUS interacting protein (serine/arginine-rich) 1	FUSIP1	604.1	-1.24514	-3.59
235117_at	similar to RIKEN cDNA 2510006C20 gene	LOC494143	2465.8	-0.35172	-3.62
1565358_at	retinoic acid receptor, alpha	RARA	3533.8	-0.22378	-3.63
209003_at	solute carrier family 25 (mitochondrial carrier; oxoglutarate carrier), member 11	SLC25A11	586.0	-1.27196	-3.69
226993_at	Thyroid hormone receptor interactor 12	TRIP12	548.8	-1.33068	-3.70
229551_x_at	zinc finger protein 367	ZNF367	3810.8	-0.20113	-3.73
222891_s_at	B-cell CLL/lymphoma 11A (zinc finger protein)	BCL11A	2611.2	-0.32899	-3.76
226556_at	Mitogen-activated protein kinase kinase kinase 13	RPL4	581.5	-1.27874	-3.85
203485_at	reticulon 1	RTN1	2529.5	-0.34150	-3.92
223339_at	ATPase inhibitory factor 1	ATPIF1	2474.1	-0.35036	-3.92
212570_at	KIAA0830 protein	KIAA0830	1086.1	-0.79024	-3.93
234987_at	Chromosome 20 open reading frame 118	C20orf118	3697.1	-0.21006	-3.93
225665_at	sterile alpha motif and leucine zipper containing kinase AZK	ZAK	717.4	-1.09965	-3.96
215029_at	---	---	2621.6	-0.32746	-3.99
228250_at	KIAA1961 gene	KIAA1961	5132.4	-0.12406	-4.00
225773_at	KIAA1972 protein	KIAA1972	2709.9	-0.31478	-4.00

230146_s_at	frequenin homolog (Drosophila)	FREQ	2097.4	-0.42110	-4.03
213357_at	general transcription factor IIH, polypeptide 5	GTF2H5	2108.4	-0.41873	-4.09
205471_s_at	dachshund homolog 1 (Drosophila)	DACH1	3840.0	-0.19891	-4.16
212249_at	phosphoinositide-3-kinase, regulatory subunit 1 (p85 alpha)	PIK3R1	1619.6	-0.54880	-4.26
218692_at	hypothetical protein FLJ20366	FLJ20366	678.6	-1.14555	-4.33
201470_at	glutathione S-transferase omega 1	GSTO1	3462.5	-0.23013	-4.34
212131_at	family with sequence similarity 61, member A	FAM61A	1926.8	-0.46065	-4.39
203255_at	F-box protein 11	FBXO11	407.4	-1.61410	-4.39
223984_s_at	nucleoporin like 1	NUPL1	2822.1	-0.29968	-4.42
222912_at	arrestin, beta 1	ARRB1	5590.9	-0.10521	-4.53
1555772_a_at	cell division cycle 25A	CDC25A	4091.4	-0.18101	-4.61
212386_at	Transcription factor 4	TCF4	2164.2	-0.40703	-4.72
205472_s_at	dachshund homolog 1 (Drosophila)	DACH1	2365.7	-0.36874	-5.01
221276_s_at	syncoilin, intermediate filament 1 /// syncoilin, intermediate filament 1	SYNC1	4188.6	-0.17461	-5.24

## APPENDIX B

<b>Class Name</b>	<b>GO ID</b>	<b>Num Probes</b>	<b>Num Genes</b>	<b>Raw Score</b>	<b>P-value</b>
DNA replication	GO:0006260	131	83	32	1.23E-05
mitochondrial electron transport, NADH to ubiquinone	GO:0006120	32	23	13	1.35E-05
mitotic checkpoint	GO:0007093	23	13	8	1.23E-04
ATP synthesis coupled electron transport (sensu Eukaryota)	GO:0042775	34	24	12	1.47E-04
neurotransmitter metabolism	GO:0042133	7	5	4	2.68E-04
mitotic spindle checkpoint	GO:0007094	9	5	4	2.68E-04
DNA replication initiation	GO:0006270	20	14	8	2.86E-04
protein amino acid dephosphorylation	GO:0006470	89	57	21	5.51E-04
cell organization and biogenesis	GO:0016043	23	15	8	5.93E-04
glutamine family amino acid metabolism	GO:0009064	14	10	6	6.87E-04
cell cycle checkpoint	GO:0000075	39	21	10	7.02E-04
M phase of mitotic cell cycle	GO:0000087	144	96	31	7.74E-04
dephosphorylation	GO:0016311	95	62	22	8.00E-04
interphase of mitotic cell cycle	GO:0051329	59	34	14	8.22E-04
positive regulation of caspase activity	GO:0043280	16	13	7	9.68E-04
traversing start control point of mitotic cell cycle	GO:0007089	10	6	4	1.35E-03
G2/M transition of mitotic cell cycle	GO:0000086	12	6	4	1.35E-03
regulation of neurotransmitter levels	GO:0001505	9	6	4	1.35E-03
transcription from RNA polymerase I promoter	GO:0006360	12	6	4	1.35E-03
DNA-dependent DNA replication	GO:0006261	52	36	14	1.67E-03
glutamine metabolism	GO:0006541	12	9	5	2.53E-03
regulation of protein kinase activity	GO:0045859	52	34	13	2.71E-03
phospholipid biosynthesis	GO:0008654	35	21	9	3.08E-03
G1 phase of mitotic cell cycle	GO:0000080	20	12	6	3.15E-03
blood vessel morphogenesis	GO:0048514	28	15	7	3.35E-03
angiogenesis	GO:0001525	27	15	7	3.35E-03
macromolecule metabolism	GO:0043170	129	90	27	4.92E-03
I-kappaB kinase/NF-kappaB cascade	GO:0007249	27	19	8	5.19E-03
iron ion transport	GO:0006826	16	10	5	5.29E-03
regulation of progression through mitotic cell cycle	GO:0007346	20	10	5	5.29E-03
cell morphogenesis	GO:0000902	28	16	7	5.57E-03
phosphoinositide-mediated signaling	GO:0048015	21	16	7	5.57E-03
positive regulation of programmed cell death	GO:0043068	31	13	6	5.69E-03

cell motility	GO:0006928	104	59	19	5.73E-03
aspartate family amino acid biosynthesis	GO:0009067	6	5	3	5.87E-03
DNA integrity checkpoint	GO:0031570	14	5	3	5.87E-03
regulation of heart contraction	GO:0008016	11	5	3	5.87E-03
physiological interaction between organisms	GO:0051706	14	5	3	5.87E-03
G-protein coupled receptor protein signaling pathway	GO:0007186	115	77	23	8.60E-03
spindle organization and biogenesis	GO:0007051	23	17	7	8.76E-03
caspase activation	GO:0006919	21	17	7	8.76E-03
localization	GO:0051179	12	8	4	8.89E-03
carbohydrate biosynthesis	GO:0016051	11	8	4	8.89E-03
regulation of cell migration	GO:0030334	18	8	4	8.89E-03
JAK-STAT cascade	GO:0007259	12	8	4	8.89E-03
G1 phase	GO:0051318	14	8	4	8.89E-03
regulation of organismal physiological process	GO:0051239	15	8	4	8.89E-03
mitotic spindle organization and biogenesis	GO:0007052	20	14	6	9.50E-03
protein localization	GO:0008104	18	11	5	9.74E-03
heart development	GO:0007507	18	11	5	9.74E-03
calcium-mediated signaling	GO:0019722	20	11	5	9.74E-03
physiological response to stimulus	GO:0051869	108	74	22	0.01034261
morphogenesis	GO:0009653	85	58	18	0.01044367
fatty acid metabolism	GO:0006631	84	58	18	0.01044367
locomotion	GO:0040011	87	51	16	0.01247124
response to stimulus	GO:0050896	139	96	27	0.01269188
regulation of mitosis	GO:0007088	43	25	9	0.0135264
apoptotic program	GO:0008632	32	25	9	0.0135264
innate immune response	GO:0045087	11	6	3	0.014907
neurotransmitter transport	GO:0006836	11	6	3	0.014907
tubulin folding	GO:0007021	10	6	3	0.014907
carbohydrate transport	GO:0008643	8	6	3	0.014907
pregnancy	GO:0007565	16	6	3	0.014907
reproductive organismal physiological process	GO:0048609	15	6	3	0.014907
N-acetylglucosamine metabolism	GO:0006044	8	6	3	0.014907
second-messenger-mediated signaling	GO:0019932	40	29	10	0.01518161
spliceosome assembly	GO:0000245	40	22	8	0.01604477
protein import into nucleus, translocation	GO:0000060	29	12	5	0.0163346
regulation of DNA replication	GO:0006275	20	12	5	0.0163346
G-protein signaling, coupled to IP3 second messenger (phospholipase C activating)	GO:0007200	17	9	4	0.01683878
carbohydrate metabolism	GO:0005975	115	77	22	0.01693955

central nervous system development	GO:0007417	37	26	9	0.01825801
cell-cell signaling	GO:0007267	72	45	14	0.01831854
negative regulation of progression through cell cycle	GO:0045786	140	86	24	0.01860027
circulation	GO:0008015	28	19	7	0.01892866
androgen receptor signaling pathway	GO:0030521	39	23	8	0.02194126
RNA splicing	GO:0008380	116	58	17	0.02208039
immune system process	GO:0002376	110	75	21	0.02364544
chromatin assembly	GO:0031497	55	27	9	0.02410063
glycogen metabolism	GO:0005977	19	13	5	0.02545936
energy reserve metabolism	GO:0006112	19	13	5	0.02545936
activation of protein kinase activity	GO:0032147	22	13	5	0.02545936
mitotic cell cycle	GO:0000278	27	20	7	0.02632545
phosphate metabolism	GO:0006796	28	20	7	0.02632545
MAPKKK cascade	GO:0000165	50	35	11	0.02653459
actin cytoskeleton organization and biogenesis	GO:0030036	84	51	15	0.02704918
single strand break repair	GO:0000012	13	10	4	0.02839203
M phase	GO:0000279	16	10	4	0.02839203
transcription initiation from RNA polymerase II promoter	GO:0006367	40	24	8	0.02924055
regulation of cyclin-dependent protein kinase activity	GO:0000079	32	24	8	0.02924055
oxygen transport	GO:0015671	12	7	3	0.02950026
regulation of apoptosis	GO:0042981	105	60	17	0.03088609
G1/S transition of mitotic cell cycle	GO:0000082	27	17	6	0.03148662
RNA metabolism	GO:0016070	142	77	21	0.03157289
ubiquitin-dependent protein catabolism	GO:0006511	154	95	25	0.03480641
immune response	GO:0006955	117	78	21	0.03621555
integrin-mediated signaling pathway	GO:0007229	26	14	5	0.03742869
protein amino acid autophosphorylation	GO:0046777	21	14	5	0.03742869
telomere maintenance	GO:0000723	32	14	5	0.03742869
protein autoprocessing	GO:0016540	22	14	5	0.03742869
modification-dependent protein catabolism	GO:0019941	133	79	21	0.04134611
JNK cascade	GO:0007254	28	18	6	0.04316049
transition metal ion transport	GO:0000041	26	18	6	0.04316049
cellular carbohydrate metabolism	GO:0044262	14	11	4	0.0439347
L-serine biosynthesis	GO:0006564	21	11	4	0.0439347
telomere organization and biogenesis	GO:0032200	27	11	4	0.0439347
nucleotide-sugar metabolism	GO:0009225	15	11	4	0.0439347
coagulation	GO:0050817	16	11	4	0.0439347
negative regulation of apoptosis	GO:0043066	116	67	18	0.04676168
lipid metabolism	GO:0006629	141	98	25	0.04939484

nucleosome assembly	GO:0006334	59	30	9	0.04944881
hydrogen transport	GO:0006818	53	34	10	0.04958215

## REFERENCES

1. Linnaeus, C. 1748. *Systema Naturae*.
2. Darwin, C. 1859. *On the Origin of Species: By Means of Natural Selection or the Preservation of Favoured Races for the Struggle of Life*. Amereon House, Mattituck, New York.
3. Lewin, R. 1999. *Complexity : life at the edge of chaos*. University of Chicago Press, Chicago, Ill. xii, 234 p. pp.
4. Kauffman, S.A. 1993. *The origins of order : self-organization and selection in evolution*. Oxford University Press, New York. xviii, 709 p. pp.
5. Descartes, R. 1637. *Discours de la methode pour bien conduire sa raison, & chercher la verité dans les sciences. Plus La dioptrique. Les meteores. Et La geometrie. Qui sont des essais de cete methode*. De l'imprimerie de I. Maire, A Leyde,. 78 p., 71 l., 413, [434] p. pp.
6. Dawkins, R. 1976. *The selfish gene*. Oxford University Press, Oxford. xii, 224 p. pp.
7. Beadle, G.W., and E.L. Tatum. 1941. Genetic Control of Biochemical Reactions in *Neurospora*. *Proc Natl Acad Sci U S A* 27:499-506.
8. Venter, J.C., M.D. Adams, E.W. Myers, P.W. Li, R.J. Mural, G.G. Sutton, H.O. Smith, M. Yandell, C.A. Evans, R.A. Holt, J.D. Gocayne, P. Amanatides, R.M. Ballew, D.H. Huson, J.R. Wortman, Q. Zhang, C.D. Kodira, X.H. Zheng, L. Chen, M. Skupski, G. Subramanian, P.D. Thomas, J. Zhang, G.L. Gabor Miklos, C. Nelson, S. Broder, A.G. Clark, J. Nadeau, V.A. McKusick, N. Zinder, A.J.

Levine, R.J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. Di Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A.E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T.J. Heiman, M.E. Higgins, R.R. Ji, Z. Ke, K.A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G.V. Merkulov, N. Milshina, H.M. Moore, A.K. Naik, V.A. Narayan, B. Neelam, D. Nusskern, D.B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Wang, A. Wang, X. Wang, J. Wang, M. Wei, R. Wides, C. Xiao, C. Yan, A. Yao, J. Ye, M. Zhan, W. Zhang, H. Zhang, Q. Zhao, L. Zheng, F. Zhong, W. Zhong, S. Zhu, S. Zhao, D. Gilbert, S. Baumhueter, G. Spier, C. Carter, A. Cravchik, T. Woodage, F. Ali, H. An, A. Awe, D. Baldwin, H. Baden, M. Barnstead, I. Barrow, K. Beeson, D. Busam, A. Carver, A. Center, M.L. Cheng, L. Curry, S. Danaher, L. Davenport, R. Desilets, S. Dietz, K. Dodson, L. Doup, S. Ferreira, N. Garg, A. Gluecksmann, B. Hart, J. Haynes, C. Haynes, C. Heiner, S. Hladun, D. Hostin, J. Houck, T. Howland, C. Ibegwam, J. Johnson, F. Kalush, L. Kline, S. Koduru, A. Love, F. Mann, D. May, S. McCawley, T. McIntosh, I. McMullen, M. Moy, L. Moy, B. Murphy, K. Nelson, C. Pfannkoch, E. Pratts, V. Puri, H. Qureshi, M. Reardon, R. Rodriguez, Y.H. Rogers, D. Romblad, B. Ruhfel, R. Scott, C. Sitter, M. Smallwood, E. Stewart, R. Strong, E. Suh, R. Thomas, N.N. Tint, S. Tse, C. Vech, G. Wang, J. Wetter, S. Williams, M.

Williams, S. Windsor, E. Winn-Deen, K. Wolfe, J. Zaveri, K. Zaveri, J.F. Abril, R. Guigo, M.J. Campbell, K.V. Sjolander, B. Karlak, A. Kejariwal, H. Mi, B. Lazareva, T. Hatton, A. Narechania, K. Diemer, A. Muruganujan, N. Guo, S. Sato, V. Bafna, S. Istrail, R. Lippert, R. Schwartz, B. Walenz, S. Yooseph, D. Allen, A. Basu, J. Baxendale, L. Blick, M. Caminha, J. Carnes-Stine, P. Caulk, Y.H. Chiang, M. Coyne, C. Dahlke, A. Mays, M. Dombroski, M. Donnelly, D. Ely, S. Esparham, C. Fosler, H. Gire, S. Glanowski, K. Glasser, A. Glodek, M. Gorokhov, K. Graham, B. Gropman, M. Harris, J. Heil, S. Henderson, J. Hoover, D. Jennings, C. Jordan, J. Jordan, J. Kasha, L. Kagan, C. Kraft, A. Levitsky, M. Lewis, X. Liu, J. Lopez, D. Ma, W. Majoros, J. McDaniel, S. Murphy, M. Newman, T. Nguyen, N. Nguyen, M. Nodell, S. Pan, J. Peck, M. Peterson, W. Rowe, R. Sanders, J. Scott, M. Simpson, T. Smith, A. Sprague, T. Stockwell, R. Turner, E. Venter, M. Wang, M. Wen, D. Wu, M. Wu, A. Xia, A. Zandieh, and X. Zhu. 2001. The sequence of the human genome. *Science* 291:1304-1351.

9. Dawkins, R. 1986. *The blind watchmaker*. Norton, New York. xiii, 332 p. pp.
10. Bartlett, J., and T.B. Backer. 1974. *A compact anthology of Bartlett's quotations*. Jonathan David Publishers, Middle Village, N.Y.,. 176 p. pp.
11. Kitano, H. 2002. Systems biology: a brief overview. *Science* 295:1662-1664.
12. Harrington, C.A., C. Rosenow, and J. Retief. 2000. Monitoring gene expression using DNA microarrays. *Curr Opin Microbiol* 3:285-291.
13. Geisow, M.J. 1998. Proteomics: one small step for a digital computer, one giant leap for humankind. *Nat Biotechnol* 16:206.

14. Anderson, N.L., and N.G. Anderson. 1998. Proteome and proteomics: new technologies, new concepts, and new words. *Electrophoresis* 19:1853-1861.
15. Nicholson, J.K., J.C. Lindon, and E. Holmes. 1999. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica* 29:1181-1189.
16. Nicholson, J.K., J. Connelly, J.C. Lindon, and E. Holmes. 2002. Metabonomics: a platform for studying drug toxicity and gene function. *Nat Rev Drug Discov* 1:153-161.
17. Nicholson, J.K., and I.D. Wilson. 2003. Opinion: understanding 'global' systems biology: metabonomics and the continuum of metabolism. *Nat Rev Drug Discov* 2:668-676.
18. Ramsay, G. 1998. DNA chips: state-of-the art. *Nat Biotechnol* 16:40-44.
19. Debouck, C., and P.N. Goodfellow. 1999. DNA microarrays in drug discovery and development. *Nat Genet* 21:48-50.
20. van Hal, N.L., O. Vorst, A.M. van Houwelingen, E.J. Kok, A. Peijnenburg, A. Aharoni, A.J. van Tunen, and J. Keijer. 2000. The application of DNA microarrays in gene expression analysis. *J Biotechnol* 78:271-280.
21. Schena, M. 1996. Genome analysis with gene expression microarrays. *Bioessays* 18:427-431.
22. Schulze, A., and J. Downward. 2000. Analysis of gene expression by microarrays: cell biologist's gold mine or minefield? *J Cell Sci* 113 Pt 23:4151-4156.

23. Herwig, R., and H. Lehrach. 2006. Expression profiling of drug response--from genes to pathways. *Dialogues Clin Neurosci* 8:283-293.
24. Kemming, D., U. Vogt, N. Tidow, C.M. Schlotter, H. Burger, M.W. Helms, E. Korsching, A. Granetzny, A. Boseila, L. Hillejan, A. Marra, Y. Ergonenc, H. Adiguzel, and B. Brandt. 2006. Whole genome expression analysis for biologic rational pathway modeling: application in cancer prognosis and therapy prediction. *Mol Diagn Ther* 10:271-280.
25. Nielsen, T.O. 2006. Microarray analysis of sarcomas. *Adv Anat Pathol* 13:166-173.
26. Hao, C.Y., D.H. Moore, P. Wong, J.L. Bennington, N.M. Lee, and L.C. Chen. 2005. Alteration of gene expression in macroscopically normal colonic mucosa from individuals with a family history of sporadic colon cancer. *Clin Cancer Res* 11:1400-1407.
27. Rowe, S.M., S. Miller, and E.J. Sorscher. 2005. Cystic fibrosis. *N Engl J Med* 352:1992-2001.
28. Steinberg, M.H., F. Barton, O. Castro, C.H. Pegelow, S.K. Ballas, A. Kutlar, E. Orringer, R. Bellevue, N. Olivieri, J. Eckman, M. Varma, G. Ramirez, B. Adler, W. Smith, T. Carlos, K. Ataga, L. DeCastro, C. Bigelow, Y. Sauntharajah, M. Telfer, E. Vichinsky, S. Claster, S. Shurin, K. Bridges, M. Waclawiw, D. Bonds, and M. Terrin. 2003. Effect of hydroxyurea on mortality and morbidity in adult sickle cell anemia: risks and benefits up to 9 years of treatment. *Jama* 289:1645-1651.

29. Yang, Z.R. 2004. Biological applications of support vector machines. *Brief Bioinform* 5:328-338.
30. Mukherjee, S., and S. Mitra. 2005. Hidden Markov Models, grammars, and biology: a tutorial. *J Bioinform Comput Biol* 3:491-526.
31. Brent, R., and J. Bruck. 2006. 2020 computing: can computers help to explain biology? *Nature* 440:416-417.
32. Muggleton, S.H. 2006. 2020 computing: exceeding human limits. *Nature* 440:409-410.
33. Kell, D.B. 2006. Theodor Bucher Lecture. Metabolomics, modelling and machine learning in systems biology - towards an understanding of the languages of cells. Delivered on 3 July 2005 at the 30th FEBS Congress and the 9th IUBMB conference in Budapest. *Febs J* 273:873-894.
34. Gianchandani, E.P., D.L. Brautigan, and J.A. Papin. 2006. Systems analyses characterize integrated functions of biochemical networks. *Trends Biochem Sci* 31:284-291.
35. Aittokallio, T., and B. Schwikowski. 2006. Graph-based methods for analysing networks in cell biology. *Brief Bioinform* 7:243-255.
36. Everett, L., L.S. Wang, and S. Hannenhalli. 2006. Dense subgraph computation via stochastic search: application to detect transcriptional modules. *Bioinformatics* 22:e117-123.
37. Lamb, J., E.D. Crawford, D. Peck, J.W. Modell, I.C. Blat, M.J. Wrobel, J. Lerner, J.P. Brunet, A. Subramanian, K.N. Ross, M. Reich, H. Hieronymus, G. Wei, S.A.

- Armstrong, S.J. Haggarty, P.A. Clemons, R. Wei, S.A. Carr, E.S. Lander, and T.R. Golub. 2006. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313:1929-1935.
38. DeRisi, J.L., V.R. Iyer, and P.O. Brown. 1997. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 278:680-686.
39. Golub, T.R., D.K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J.P. Mesirov, H. Coller, M.L. Loh, J.R. Downing, M.A. Caligiuri, C.D. Bloomfield, and E.S. Lander. 1999. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286:531-537.
40. Hughes, T.R., M.J. Marton, A.R. Jones, C.J. Roberts, R. Stoughton, C.D. Armour, H.A. Bennett, E. Coffey, H. Dai, Y.D. He, M.J. Kidd, A.M. King, M.R. Meyer, D. Slade, P.Y. Lum, S.B. Stepaniants, D.D. Shoemaker, D. Gachotte, K. Chakraborty, J. Simon, M. Bard, and S.H. Friend. 2000. Functional discovery via a compendium of expression profiles. *Cell* 102:109-126.
41. Hastie, T., R. Tibshirani, and J.H. Friedman. 2001. The elements of statistical learning : data mining, inference, and prediction : with 200 full-color illustrations. Springer, New York. xvi, 533 p. pp.
42. Jiang, Z., and Y. Zhou. 2005. Using bioinformatics for drug target identification from the genome. *Am J Pharmacogenomics* 5:387-396.
43. Pal, C., B. Papp, and M.J. Lercher. 2006. An integrated view of protein evolution. *Nat Rev Genet* 7:337-348.

44. Overall, C.M., and R.A. Dean. 2006. Degradomics: systems biology of the protease web. Pleiotropic roles of MMPs in cancer. *Cancer Metastasis Rev* 25:69-75.
45. Ackermann, B.L., J.E. Hale, and K.L. Duffin. 2006. The role of mass spectrometry in biomarker discovery and measurement. *Curr Drug Metab* 7:525-539.
46. Ren, B., F. Robert, J.J. Wyrick, O. Aparicio, E.G. Jennings, I. Simon, J. Zeitlinger, J. Schreiber, N. Hannett, E. Kanin, T.L. Volkert, C.J. Wilson, S.P. Bell, and R.A. Young. 2000. Genome-wide location and function of DNA binding proteins. *Science* 290:2306-2309.
47. Lee, T.I., N.J. Rinaldi, F. Robert, D.T. Odom, Z. Bar-Joseph, G.K. Gerber, N.M. Hannett, C.T. Harbison, C.M. Thompson, I. Simon, J. Zeitlinger, E.G. Jennings, H.L. Murray, D.B. Gordon, B. Ren, J.J. Wyrick, J.B. Tagne, T.L. Volkert, E. Fraenkel, D.K. Gifford, and R.A. Young. 2002. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* 298:799-804.
48. Ren, B., and B.D. Dynlacht. 2004. Use of chromatin immunoprecipitation assays in genome-wide location analysis of mammalian transcription factors. *Methods Enzymol* 376:304-315.
49. Husmeier, D. 2003. Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic Bayesian networks. *Bioinformatics* 19:2271-2282.

50. Yu, J., V.A. Smith, P.P. Wang, A.J. Hartemink, and E.D. Jarvis. 2004. Advances to Bayesian network inference for generating causal networks from observational biological data. *Bioinformatics* 20:3594-3603.
51. Bernard, A., and A.J. Hartemink. 2005. Informative structure priors: joint learning of dynamic regulatory networks from multiple types of data. *Pac Symp Biocomput* 459-470.
52. Schafer, J., and K. Strimmer. 2005. An empirical Bayes approach to inferring large-scale gene association networks. *Bioinformatics* 21:754-764.
53. Ingenuity. Curated online systems biology resource for biomolecular pathway inference.
54. Si, W., Q. Kang, H.H. Luu, J.K. Park, Q. Luo, W.X. Song, W. Jiang, X. Luo, X. Li, H. Yin, A.G. Montag, R.C. Haydon, and T.C. He. 2006. CCN1/Cyr61 is regulated by the canonical Wnt signal and plays an important role in Wnt3A-induced osteoblast differentiation of mesenchymal stem cells. *Mol Cell Biol* 26:2955-2964.
55. Song, L., N.E. Webb, Y. Song, and R.S. Tuan. 2006. Identification and functional analysis of candidate genes regulating mesenchymal stem cell self-renewal and multipotency. *Stem Cells* 24:1707-1718.
56. Zhou, J., E.F. Rappaport, J.W. Tobias, and T.L. Young. 2006. Differential gene expression in mouse sclera during ocular development. *Invest Ophthalmol Vis Sci* 47:1794-1802.

57. Maruyama, R., F. Aoki, M. Toyota, Y. Sasaki, H. Akashi, H. Mita, H. Suzuki, K. Akino, M. Ohe-Toyota, Y. Maruyama, H. Tatsumi, K. Imai, Y. Shinomura, and T. Tokino. 2006. Comparative genome analysis identifies the vitamin D receptor gene as a direct target of p53-mediated transcriptional activation. *Cancer Res* 66:4574-4583.
58. Luthra, R., T.T. Wu, M.G. Luthra, J. Izzo, E. Lopez-Alvarez, L. Zhang, J. Bailey, J.H. Lee, R. Bresalier, A. Rashid, S.G. Swisher, and J.A. Ajani. 2006. Gene expression profiling of localized esophageal carcinomas: association with pathologic response to preoperative chemoradiation. *J Clin Oncol* 24:259-267.
59. Bredel, M., C. Bredel, D. Juric, G.E. Duran, R.X. Yu, G.R. Harsh, H. Vogel, L.D. Recht, A.C. Scheck, and B.I. Sikic. 2006. Tumor necrosis factor-alpha-induced protein 3 as a putative regulator of nuclear factor-kappaB-mediated resistance to O6-alkylating agents in human glioblastomas. *J Clin Oncol* 24:274-287.
60. Grade, M., B.M. Ghadimi, S. Varma, R. Simon, D. Wangsa, L. Barenboim-Stapleton, T. Liersch, H. Becker, T. Ried, and M.J. Difilippantonio. 2006. Aneuploidy-dependent massive deregulation of the cellular transcriptome and apparent divergence of the Wnt/beta-catenin signaling pathway in human rectal carcinomas. *Cancer Res* 66:267-282.
61. Juric, D., S. Sale, R.A. Hromas, R. Yu, Y. Wang, G.E. Duran, R. Tibshirani, L.H. Einhorn, and B.I. Sikic. 2005. Gene expression profiling differentiates germ cell tumors from other cancers and defines subtype-specific signatures. *Proc Natl Acad Sci U S A* 102:17763-17768.

62. Gielen, S.C., L.C. Kuhne, P.C. Ewing, L.J. Blok, and C.W. Burger. 2005. Tamoxifen treatment for breast cancer enforces a distinct gene-expression profile on the human endometrium: an exploratory study. *Endocr Relat Cancer* 12:1037-1049.
63. Liu, Y., L. Chen, T.C. Ko, A.P. Fields, and E.A. Thompson. 2006. Evi1 is a survival factor which conveys resistance to both TGFbeta- and taxol-mediated cell death via PI3K/AKT. *Oncogene* 25:3565-3575.
64. Murakami, M., K. Shimada, M. Kawai, and H. Koga. 2005. InCeP: Intracellular Pathway Based on mKIAA Protein-Protein Interactions. *DNA Res* 12:379-387.
65. Gerling, I.C., S. Singh, N.I. Lenchik, D.R. Marshall, and J. Wu. 2006. New data analysis and mining approaches identify unique proteome and transcriptome markers of susceptibility to autoimmune diabetes. *Mol Cell Proteomics* 5:293-305.
66. Waters, K.M., J.G. Pounds, and B.D. Thrall. 2006. Data merging for integrated microarray and proteomic analysis. *Brief Funct Genomic Proteomic*
67. Grigorov, M.G. 2006. Global dynamics of biological systems from time-resolved omics experiments. *Bioinformatics* 22:1424-1430.
68. Huang, Y., J. Yan, R. Lubet, T.W. Kensler, and T.R. Sutter. 2006. Identification of novel transcriptional networks in response to treatment with the anticarcinogen 3H-1,2-dithiole-3-thione. *Physiol Genomics* 24:144-153.

69. Irizarry, R.A., B. Hobbs, F. Collin, Y.D. Beazer-Barclay, K.J. Antonellis, U. Scherf, and T.P. Speed. 2003. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4:249-264.
70. Li, C., and W.H. Wong. 2001. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc Natl Acad Sci U S A* 98:31-36.
71. Wu, Z., Irizarry, R.A., Gentleman, R., Murillo, F.M., Spencer, F. 2004. A Model Based Background Adjustment for Oligonucleotide Expression Arrays. In Dept. of Biostatistics Working Papers, Working Paper 1. Johns Hopkins University, Baltimore.
72. Irizarry, R.A., Z. Wu, and H.A. Jaffee. 2006. Comparison of Affymetrix GeneChip expression measures. *Bioinformatics* 22:789-794.
73. Mangelsdorf, D.J., and R.M. Evans. 1995. The RXR heterodimers and orphan receptors. *Cell* 83:841-850.
74. Berg, J.M. 1989. DNA binding specificity of steroid receptors. *Cell* 57:1065-1068.
75. Klug, A., and J.W. Schwabe. 1995. Protein motifs 5. Zinc fingers. *Faseb J* 9:597-604.
76. Bourguet, W., P. Germain, and H. Gronemeyer. 2000. Nuclear receptor ligand-binding domains: three-dimensional structures, molecular interactions and pharmacological implications. 21:381-388.

77. Aranda, A., and A. Pascual. 2001. Nuclear Hormone Receptors and Gene Expression. *Physiol.Rev.* 81:1269-1304.
78. Issemann, I., and S. Green. 1990. Activation of a member of the steroid hormone receptor superfamily by peroxisome proliferators. 347:645-650.
79. Chen, L. 2003. The role of PPARgamma in the development of normal intestinal epithelium. In Human Biological Chemistry and Genetics. University of Texas Medical Branch, Galveston. 213.
80. Dreyer, C., G. Krey, H. Keller, F. Givel, G. Helftenbein, and W. Wahli. 1992. Control of the peroxisomal beta-oxidation pathway by a novel family of nuclear hormone receptors. *Cell* 68:879-887.
81. Zhu, Y., K. Alvares, Q. Huang, M.S. Rao, and J.K. Reddy. 1993. Cloning of a new member of the peroxisome proliferator-activated receptor gene family from mouse liver. *J Biol Chem.* 268:26817-26820.
82. Kliewer, S.A., B.M. Forman, B. Blumberg, E.S. Ong, U. Borgmeyer, D.J. Mangelsdorf, K. Umesono, and R.M. Evans. 1994. Differential expression and activation of a family of murine peroxisome proliferator-activated receptors. *Proc Natl Acad Sci U S A* 91:7355-7359.
83. Xing, G., L. Zhang, L. Zhang, T. Heynen, T. Yoshikawa, M. Smith, S. Weiss, and S. Detera-Wadleigh. 1995. Rat PPAR delta contains a CGG triplet repeat and is prominently expressed in the thalamic nuclei. *Biochem.Biophys.Res.Commun.* 217:1015-1025.

84. Aperlo, C., P. Pognonec, R. Saladin, J. Auwerx, and K.E. Boulukos. 1995. cDNA cloning and characterization of the transcriptional activities of the hamster peroxisome proliferator-activated receptor haPPAR gamma. *Gene* 162:297-302.
85. Sher, T., H.F. Yi, O.W. McBride, and F.J. Gonzalez. 1993. cDNA cloning, chromosomal mapping, and functional characterization of the human peroxisome proliferator activated receptor. *Biochemistry* 32:5598-5604.
86. Greene, M.E., B. Blumberg, O.W. McBride, H.F. Yi, K. Kronquist, K. Kwan, L. Hsieh, G. Greene, and S.D. Nimer. 1995. Isolation of the human peroxisome proliferator activated receptor gamma cDNA: expression in hematopoietic cells and chromosomal mapping. *Gene Expr* 4:281-299.
87. Tugwood, J.D., I. Issemann, R.G. Anderson, K.R. Bundell, W.L. McPheat, and S. Green. 1992. The mouse peroxisome proliferator activated receptor recognizes a response element in the 5' flanking sequence of the rat acyl CoA oxidase gene. *Embo J* 11:433-439.
88. Kadera, Y., K. Takeyama, A. Murayama, M. Suzawa, Y. Masuhiro, and S. Kato. 2000. Ligand type-specific interactions of peroxisome proliferator-activated receptor gamma with transcriptional coactivators. *J Biol Chem.* 275:33201-33204.
89. Zhu, Y., C. Qi, J.R. Korenberg, X.N. Chen, D. Noya, M.S. Rao, and J.K. Reddy. 1995. Structural organization of mouse peroxisome proliferator-activated receptor gamma (mPPAR gamma) gene: alternative promoter use and different splicing yield two mPPAR gamma isoforms. *Proc Natl Acad Sci U S A* 92:7921-7925.

90. Mukherjee, R., L. Jow, G.E. Croston, and J.R. Paterniti, Jr. 1997. Identification, characterization, and tissue distribution of human peroxisome proliferator-activated receptor (PPAR) isoforms PPARgamma2 versus PPARgamma1 and activation with retinoid X receptor agonists and antagonists. *J Biol Chem* 272:8071-8076.
91. Fajas, L., D. Auboeuf, E. Raspe, K. Schoonjans, A.M. Lefebvre, R. Saladin, J. Najib, M. Laville, J.C. Fruchart, S. Deeb, A. Vidal-Puig, J. Flier, M.R. Briggs, B. Staels, H. Vidal, and J. Auwerx. 1997. The organization, promoter analysis, and expression of the human PPARgamma gene. *J Biol Chem* 272:18779-18789.
92. Fajas, L., J.C. Fruchart, and J. Auwerx. 1998. PPARgamma3 mRNA: a distinct PPARgamma mRNA subtype transcribed from an independent promoter. *FEBS Lett.* 438:55-60.
93. Ricote, M., J. Huang, L. Fajas, A. Li, J. Welch, J. Najib, J.L. Witztum, J. Auwerx, W. Palinski, and C.K. Glass. 1998. Expression of the peroxisome proliferator-activated receptor gamma (PPARgamma) in human atherosclerosis and regulation in macrophages by colony stimulating factors and oxidized low density lipoprotein. *Proc.Natl.Acad.Sci U.S.A* 95:7614-7619.
94. Motojima, K. 1993. Peroxisome proliferator-activated receptor (PPAR): structure, mechanisms of activation and diverse functions. *Cell Struct.Funct.* 18:267-277.
95. He, T.C., T.A. Chan, B. Vogelstein, and K.W. Kinzler. 1999. PPARdelta is an APC-regulated target of nonsteroidal anti-inflammatory drugs. *Cell* 99:335-345.

96. Xu, H.E., M.H. Lambert, V.G. Montana, D.J. Parks, S.G. Blanchard, P.J. Brown, D.D. Sternbach, J.M. Lehmann, G.B. Wisely, T.M. Willson, S.A. Kliewer, and M.V. Milburn. 1999. Molecular recognition of fatty acids by peroxisome proliferator-activated receptors. *Mol Cell* 3:397-403.
97. Kliewer, S.A., J.M. Lenhard, T.M. Willson, I. Patel, D.C. Morris, and J.M. Lehmann. 1995. A prostaglandin J2 metabolite binds peroxisome proliferator-activated receptor gamma and promotes adipocyte differentiation. *Cell* 83:813-819.
98. Forman, B.M., P. Tontonoz, J. Chen, R.P. Brun, B.M. Spiegelman, and R.M. Evans. 1995. 15-Deoxy-delta 12, 14-prostaglandin J2 is a ligand for the adipocyte determination factor PPAR gamma. *Cell* 83:803-812.
99. Davies, S.S., A.V. Pontsler, G.K. Marathe, K.A. Harrison, R.C. Murphy, J.C. Hinshaw, G.D. Prestwich, A.S. Hilaire, S.M. Prescott, G.A. Zimmerman, and T.M. McIntyre. 2001. Oxidized alkyl phospholipids are specific, high affinity peroxisome proliferator-activated receptor gamma ligands and agonists. *J Biol Chem*. 276:16015-16023.
100. Fujiwara, T., S. Yoshioka, T. Yoshioka, I. Ushiyama, and H. Horikoshi. 1988. Characterization of new oral antidiabetic agent CS-045. Studies in KK and ob/ob mice and Zucker fatty rats. *Diabetes* 37:1549-1558.
101. Saltiel, A.R., and J.M. Olefsky. 1996. Thiazolidinediones in the treatment of insulin resistance and type II diabetes. *Diabetes* 45:1661-1669.

102. Lehmann, J.M., L.B. Moore, T.A. Smith-Oliver, W.O. Wilkison, T.M. Willson, and S.A. Kliewer. 1995. An Antidiabetic Thiazolidinedione Is a High Affinity Ligand for Peroxisome Proliferator-activated Receptor [IMAGE] (PPAR[IMAGE]). *J. Biol. Chem.* 270:12953-12956.
103. Willson, T.M., J.E. Cobb, D.J. Cowan, R.W. Wiethe, I.D. Correa, S.R. Prakash, K.D. Beck, L.B. Moore, S.A. Kliewer, and J.M. Lehmann. 1996. The structure-activity relationship between peroxisome proliferator-activated receptor gamma agonism and the antihyperglycemic activity of thiazolidinediones. *J Med.Chem.* 39:665-668.
104. Zhu, X., Y. Lin, J. Zhang, M. Fu, Z. Mao, and Y.E. Chen. 2003. Thiazolidinediones, a class of anti-diabetic drugs, inhibit Id2 expression through a PPARgamma-independent pathway in human aortic smooth muscle cells. *Cell Mol Life Sci* 60:212-218.
105. Wang, M., S.C. Wise, T. Leff, and T.Z. Su. 1999. Troglitazone, an antidiabetic agent, inhibits cholesterol biosynthesis through a mechanism independent of peroxisome proliferator-activated receptor-gamma. *Diabetes* 48:254-260.
106. Brunmair, B., F. Gras, S. Neschen, M. Roden, L. Wagner, W. Waldhausl, and C. Furnsinn. 2001. Direct Thiazolidinedione Action on Isolated Rat Skeletal Muscle Fuel Handling Is Independent of Peroxisome Proliferator-Activated Receptor- $\{\gamma\}$ -Mediated Changes in Gene Expression. *Diabetes* 50:2309-2315.
107. Baek, S.J., L.C. Wilson, L.C. Hsi, and T.E. Eling. 2003. Troglitazone, a Peroxisome Proliferator-activated Receptor gamma (PPARgamma ) Ligand,

- Selectively Induces the Early Growth Response-1 Gene Independently of PPARgamma . A NOVEL MECHANISM FOR ITS ANTI-TUMORIGENIC ACTIVITY. *J.Biol.Chem.* 278:5845-5853.
108. Fujita, T., T. Fujiwara, and T. Izumi. 2004. Thiazolidine-2,4-dione hydrochloride salt, pharmaceutical compositions thereof and treatment method therewith. In USPTO, editor Sankyo Company, Limited (Tokyo, JP) Japan.
  109. Chen, L., C.R. Bush, B.M. Necela, W. Su, M. Yanagisawa, P.Z. Anastasiadis, A.P. Fields, and E.A. Thompson. 2006. RS5444, a novel PPARgamma agonist, regulates aspects of the differentiated phenotype in nontransformed intestinal epithelial cells. *Mol Cell Endocrinol*
  110. American-Cancer-Society. 2006. Cancer Facts and Figures.
  111. Lefebvre, M., B. Paulweber, L. Fajas, J. Woods, C. McCrary, J.F. Colombel, J. Najib, J.C. Fruchart, C. Datz, H. Vidal, P. Desreumaux, and J. Auwerx. 1999. Peroxisome proliferator-activated receptor gamma is induced during differentiation of colon epithelium cells. *J Endocrinol* 162:331-340.
  112. Sarraf, P., E. Mueller, D. Jones, F.J. King, D.J. DeAngelo, J.B. Partridge, S.A. Holden, L.B. Chen, S. Singer, C. Fletcher, and B.M. Spiegelman. 1998. Differentiation and reversal of malignant changes in colon cancer through PPARgamma. *Nat Med* 4:1046-1052.
  113. Kitamura, S., Y. Miyazaki, Y. Shinomura, S. Kondo, S. Kanayama, and Y. Matsuzawa. 1999. Peroxisome proliferator-activated receptor gamma induces

- growth arrest and differentiation markers of human colon cancer cells. *Jpn J Cancer Res* 90:75-80.
114. Tanaka, T., H. Kohno, S. Yoshitani, S. Takashima, A. Okumura, A. Murakami, and M. Hosokawa. 2001. Ligands for peroxisome proliferator-activated receptors alpha and gamma inhibit chemically induced colitis and formation of aberrant crypt foci in rats. *Cancer Res.* 61:2424-2428.
  115. Sarraf, P., E. Mueller, W.M. Smith, H.M. Wright, J.B. Kum, L.A. Aaltonen, A. de la Chapelle, B.M. Spiegelman, and C. Eng. 1999. Loss-of-function mutations in PPAR gamma associated with human colon cancer. *Molecular Cell* 3:799-804.
  116. Kinzler, K.W., and B. Vogelstein. 1996. Lessons from hereditary colorectal cancer. *Cell* 87:159-170.
  117. Jacoby, R.F., D.J. Marshall, M.A. Newton, K. Novakovic, K. Tutsch, C.E. Cole, R.A. Lubet, G.J. Kelloff, A. Verma, A.R. Moser, and W.F. Dove. 1996. Chemoprevention of spontaneous intestinal adenomas in the Apc Min mouse model by the nonsteroidal anti-inflammatory drug piroxicam. *Cancer Res.* 56:710-714.
  118. Oshima, M., J.E. Dinchuk, S.L. Kargman, H. Oshima, B. Hancock, E. Kwong, J.M. Trzaskos, J.F. Evans, and M.M. Taketo. 1996. Suppression of intestinal polyposis in Apc delta716 knockout mice by inhibition of cyclooxygenase 2 (COX-2). *Cell* 87:803-809.
  119. Lefebvre, A.M., I. Chen, P. Desreumaux, J. Najib, J.C. Fruchart, K. Geboes, M. Briggs, R. Heyman, and J. Auwerx. 1998. Activation of the peroxisome

- proliferator-activated receptor gamma promotes the development of colon tumors in C57BL/6J-APC<sup>Min/+</sup> mice. *Nat Med* 4:1053-1057.
120. Saez, E., P. Tontonoz, M.C. Nelson, J.G. Alvarez, U.T. Ming, S.M. Baird, V.A. Thomazy, and R.M. Evans. 1998. Activators of the nuclear receptor PPAR $\gamma$  enhance colon polyp formation. *Nat Med* 4:1058-1061.
  121. Pino, M.V., M.F. Kelley, and Z. Jayyosi. 2004. Promotion of colon tumors in C57BL/6J-APC(min)/+ mice by thiazolidinedione PPAR $\gamma$  agonists and a structurally unrelated PPAR $\gamma$  agonist. *Toxicol Pathol* 32:58-63.
  122. Yang, K., K.H. Fan, S.A. Lamprecht, W. Edelmann, L. Kopelovich, R. Kucherlapati, and M. Lipkin. 2005. Peroxisome proliferator-activated receptor gamma agonist troglitazone induces colon tumors in normal C57BL/6J mice and enhances colonic carcinogenesis in Apc<sup>1638 N/+</sup> Mlh1<sup>+/-</sup> double mutant mice. *Int J Cancer* 116:495-499.
  123. McAlpine, C.A., Y. Barak, I. Matisse, and R.T. Cormier. 2006. Intestinal-specific PPAR $\gamma$  deficiency enhances tumorigenesis in Apc(Min/+) mice. *International journal of cancer* 119:2339-2346.
  124. Su, L.K., K.W. Kinzler, B. Vogelstein, A.C. Preisinger, A.R. Moser, C. Luongo, K.A. Gould, and W.F. Dove. 1992. Multiple intestinal neoplasia caused by a mutation in the murine homolog of the APC gene. *Science* 256:668-670.
  125. Powell, S.M., N. Zilz, Y. Beazer-Barclay, T.M. Bryan, S.R. Hamilton, S.N. Thibodeau, B. Vogelstein, and K.W. Kinzler. 1992. APC mutations occur early during colorectal tumorigenesis. 359:235-237.

126. Girnun, G.D., W.M. Smith, S. Drori, P. Sarraf, E. Mueller, C. Eng, P. Nambiar, D.W. Rosenberg, R.T. Bronson, W. Edelmann, R. Kucherlapati, F.J. Gonzalez, and B.M. Spiegelman. 2002. APC-dependent suppression of colon carcinogenesis by PPARgamma *PNAS* 99:13771-13776.
127. Fearon, E.R., and B. Vogelstein. 1990. A genetic model for colorectal tumorigenesis. *Cell* 61:759-767.
128. Hu, E., J.B. Kim, P. Sarraf, and B.M. Spiegelman. 1996. Inhibition of adipogenesis through MAP kinase-mediated phosphorylation of PPARgamma. *Science* 274:2100-2103.
129. Adams, M., M.J. Reginato, D. Shao, M.A. Lazar, and V.K. Chatterjee. 1997. Transcriptional activation by peroxisome proliferator-activated receptor gamma is inhibited by phosphorylation at a consensus mitogen-activated protein kinase site. *J Biol Chem* 272:5128-5132.
130. Antonarakis, S.E. 1998. 10 years of Genomics, chromosome 21, and Down syndrome. *Genomics* 51:1-16.
131. Fuentes, J.J., M.A. Pritchard, A.M. Planas, A. Bosch, I. Ferrer, and X. Estivill. 1995. A new human gene from the Down syndrome critical region encodes a proline-rich protein highly expressed in fetal brain and heart. *Hum Mol Genet* 4:1935-1944.
132. Fuentes, J.J., M.A. Pritchard, and X. Estivill. 1997. Genomic organization, alternative splicing, and expression patterns of the DSCR1 (Down syndrome candidate region 1) gene. *Genomics* 44:358-361.

133. Rothermel, B., R.B. Vega, J. Yang, H. Wu, R. Bassel-Duby, and R.S. Williams. 2000. A protein encoded within the Down syndrome critical region is enriched in striated muscles and inhibits calcineurin signaling. *J Biol Chem* 275:8719-8725.
134. Ermak, G., C.D. Harris, and K.J. Davies. 2002. The DSCR1 (Adapt78) isoform 1 protein calcipressin 1 inhibits calcineurin and protects against acute calcium-mediated stress damage, including transient oxidative stress. *Faseb J* 16:814-824.
135. Lin, H.Y., H.J. Michtalik, S. Zhang, T.T. Andersen, D.A. Van Riper, K.K. Davies, G. Ermak, L.M. Petti, S. Nachod, A.V. Narayan, N. Bhatt, and D.R. Crawford. 2003. Oxidative and calcium stress regulate DSCR1 (Adapt78/MCIP1) protein. *Free Radic Biol Med* 35:528-539.
136. Yao, Y.G., and E.J. Duh. 2004. VEGF selectively induces Down syndrome critical region 1 gene expression in endothelial cells: a mechanism for feedback regulation of angiogenesis? *Biochem Biophys Res Commun* 321:648-656.
137. Hesser, B.A., X.H. Liang, G. Camenisch, S. Yang, D.A. Lewin, R. Scheller, N. Ferrara, and H.P. Gerber. 2004. Down syndrome critical region protein 1 (DSCR1), a novel VEGF target gene that regulates expression of inflammatory markers on activated endothelial cells. *Blood* 104:149-158.
138. Hampton, T. 2005. Down syndrome protein deters cancer: scientists reveal molecular mechanism. *Jama* 293:284-285.
139. Minami, T., K. Horiuchi, M. Miura, M.R. Abid, W. Takabe, N. Noguchi, T. Kohro, X. Ge, H. Aburatani, T. Hamakubo, T. Kodama, and W.C. Aird. 2004. Vascular endothelial growth factor- and thrombin-induced termination factor,

- Down syndrome critical region-1, attenuates endothelial cell proliferation and angiogenesis. *J Biol Chem* 279:50537-50554.
140. Clipstone, N.A., and G.R. Crabtree. 1992. Identification of calcineurin as a key signalling enzyme in T-lymphocyte activation. *Nature* 357:695-697.
  141. Crabtree, G.R., and E.N. Olson. 2002. NFAT signaling: choreographing the social lives of cells. *Cell* 109 Suppl:S67-79.
  142. Jauliac, S., C. Lopez-Rodriguez, L.M. Shaw, L.F. Brown, A. Rao, and A. Toker. 2002. The role of NFAT transcription factors in integrin-mediated carcinoma invasion. *Nat Cell Biol* 4:540-544.
  143. Yoeli-Lerner, M., G.K. Yiu, I. Rabinovitz, P. Erhardt, S. Jauliac, and A. Toker. 2005. Akt blocks breast cancer cell motility and invasion through the transcription factor NFAT. *Mol Cell* 20:539-550.
  144. Yiu, G.K., and A. Toker. 2006. NFAT induces breast cancer cell invasion by promoting the induction of cyclooxygenase-2. *J Biol Chem* 281:12210-12217.
  145. Klee, C.B., H. Ren, and X. Wang. 1998. Regulation of the calmodulin-stimulated protein phosphatase, calcineurin. *J Biol Chem* 273:13367-13370.
  146. Kingsbury, T.J., and K.W. Cunningham. 2000. A conserved family of calcineurin regulators. *Genes Dev* 14:1595-1604.
  147. Rothermel, B.A., R.B. Vega, and R.S. Williams. 2003. The role of modulatory calcineurin-interacting proteins in calcineurin signaling. *Trends Cardiovasc Med* 13:15-21.

148. Harris, C.D., G. Ermak, and K.J. Davies. 2005. Multiple roles of the DSCR1 (Adapt78 or RCAN1) gene and its protein product calcipressin 1 (or RCAN1) in disease. *Cell Mol Life Sci* 62:2477-2486.
149. Vega, R.B., J. Yang, B.A. Rothermel, R. Bassel-Duby, and R.S. Williams. 2002. Multiple domains of MCIP1 contribute to inhibition of calcineurin activity. *J Biol Chem* 277:30401-30407.
150. Chan, B., G. Greenan, F. McKeon, and T. Ellenberger. 2005. Identification of a peptide fragment of DSCR1 that competitively inhibits calcineurin activity in vitro and in vivo. *Proc Natl Acad Sci U S A* 102:13075-13080.
151. Rao, A., C. Luo, and P.G. Hogan. 1997. Transcription factors of the NFAT family: regulation and function. *Annu Rev Immunol* 15:707-747.
152. Stathatos, N., I. Bourdeau, A.V. Espinosa, M. Saji, V.V. Vasko, K.D. Burman, C.A. Stratakis, and M.D. Ringel. 2005. KiSS-1/G protein-coupled receptor 54 metastasis suppressor pathway increases myocyte-enriched calcineurin interacting protein 1 expression and chronically inhibits calcineurin activity. *J Clin Endocrinol Metab* 90:5432-5440.
153. Duque, J., M. Fresno, and M.A. Iniguez. 2005. Expression and function of the nuclear factor of activated T cells in colon carcinoma cells: involvement in the regulation of cyclooxygenase-2. *J Biol Chem* 280:8686-8693.
154. Eberhart, C.E., R.J. Coffey, A. Radhika, F.M. Giardiello, S. Ferrenbach, and R.N. DuBois. 1994. Up-regulation of cyclooxygenase 2 gene expression in human colorectal adenomas and adenocarcinomas. *Gastroenterology* 107:1183-1188.

155. Tsujii, M., S. Kawano, and R.N. DuBois. 1997. Cyclooxygenase-2 expression in human colon cancer cells increases metastatic potential. *Proc Natl Acad Sci U S A* 94:3336-3340.
156. Buchholz, M., A. Schatz, M. Wagner, P. Michl, T. Linhart, G. Adler, T.M. Gress, and V. Ellenrieder. 2006. Overexpression of c-myc in pancreatic cancer caused by ectopic activation of NFATc1 and the Ca<sup>2+</sup>/calcineurin signaling pathway. *Embo J* 25:3714-3724.
157. Lakshmikuttyamma, A., P. Selvakumar, R. Kanthan, S.C. Kanthan, and R.K. Sharma. 2005. Increased expression of calcineurin in human colorectal adenocarcinomas. *J Cell Biochem* 95:731-739.
158. Lehmann, J.M., L.B. Moore, T.A. Smith-Oliver, W.O. Wilkison, T.M. Willson, and S.A. Kliewer. 1995. An antidiabetic thiazolidinedione is a high affinity ligand for peroxisome proliferator-activated receptor gamma (PPAR gamma). *J Biol Chem*. 270:12953-12956.
159. Kliewer, S.A., K. Umesono, D.J. Noonan, R.A. Heyman, and R.M. Evans. 1992. Convergence of 9-cis retinoic acid and peroxisome proliferator signalling pathways through heterodimer formation of their receptors. 358:771-774.
160. Technelysium. Chromas Lite. In Technelysium Pty Ltd. Copyright 1998-2005.
161. Bradford, M.M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem*. 72:248-254.

162. Knoll, T. 2002. Adobe Photoshop. In Adobe Systems Incorporated Copyright 1990-2002.
163. Mathworks. 2005. MATLAB. In The Mathworks Incorporated Copyright 1984-2005, Language for technical computing.
164. Livak, K.J., and T.D. Schmittgen. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta C(T)}$  Method. *Methods* 25:402-408.
165. Pratt, W., and P. Taylor. 1990. Principles of Drug Action: The Basis of Pharmacology. Churchill Livingstone, Ann Arbor, Michigan.
166. Team, R.D.C. 2006. R: A Language and Environment for Statistical Computing. Vienna, Austria.
167. Gentleman, R.C., V.J. Carey, D.M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. Irizarry, F. Leisch, C. Li, M. Maechler, A.J. Rossini, G. Sawitzki, C. Smith, G. Smyth, L. Tierney, J.Y. Yang, and J. Zhang. 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 5:R80.
168. Dudoit, S., R.C. Gentleman, and J. Quackenbush. 2003. Open source software for the analysis of microarray data. *Biotechniques* Suppl:45-51.
169. O'Connell, M. 2003. Differential expression, class discovery and class prediction using S-Plus and S+Array Analyzer. *SIGKDD Exploration* 5:

170. Dudoit, S., J. Shaffer, and J. Boldrick. 2002. Multiple Hypothesis Testing in Microarray Experiments. *U.C. Berkeley Division of Biostatistics Working Paper Series*
171. Conesa, A., M.J. Nueda, A. Ferrer, and M. Talon. 2006. maSigPro: a method to identify significantly differential expression profiles in time-course microarray experiments. *Bioinformatics* 22:1096-1102.
172. Reiner, A., D. Yekutieli, and Y. Benjamini. 2003. Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics* 19:368-375.
173. Draper, N.R., and H. Smith. 1998. Applied regression analysis. Wiley, New York. xvii, 706 p. pp.
174. Tai, Y.C., and T.C. Speed. 2004. A multivariate empirical Bayes statistic for replicated microarray time course data. *Technical report 667, Department of Statistics, Johns Hopkins University*
175. Tai, Y.C., and T.C. Speed. 2005. Longitudinal Microarray Time Course MB-statistic for Multiple Sample Groups. *Department of Statistics, University of California, Berkeley (In preparation)*
176. Ashburner, M., C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J.C. Matese, J.E. Richardson, M. Ringwald, G.M. Rubin, and G. Sherlock. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25:25-29.

177. Harris, M.A., J. Clark, A. Ireland, J. Lomax, M. Ashburner, R. Foulger, K. Eilbeck, S. Lewis, B. Marshall, C. Mungall, J. Richter, G.M. Rubin, J.A. Blake, C. Bult, M. Dolan, H. Drabkin, J.T. Eppig, D.P. Hill, L. Ni, M. Ringwald, R. Balakrishnan, J.M. Cherry, K.R. Christie, M.C. Costanzo, S.S. Dwight, S. Engel, D.G. Fisk, J.E. Hirschman, E.L. Hong, R.S. Nash, A. Sethuraman, C.L. Theesfeld, D. Botstein, K. Dolinski, B. Feierbach, T. Berardini, S. Mundodi, S.Y. Rhee, R. Apweiler, D. Barrell, E. Camon, E. Dimmer, V. Lee, R. Chisholm, P. Gaudet, W. Kibbe, R. Kishore, E.M. Schwarz, P. Sternberg, M. Gwinn, L. Hannick, J. Wortman, M. Berriman, V. Wood, N. de la Cruz, P. Tonellato, P. Jaiswal, T. Seigfried, and R. White. 2004. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* 32:D258-261.
178. Pavlidis, P., J. Qin, V. Arango, J.J. Mann, and E. Sibille. 2004. Using the gene ontology for microarray data mining: a comparison of methods and application to age effects in human prefrontal cortex. *Neurochem Res* 29:1213-1222.
179. Way, J.M., W.W. Harrington, K.K. Brown, W.K. Gottschalk, S.S. Sundseth, T.A. Mansfield, R.K. Ramachandran, T.M. Willson, and S.A. Kliewer. 2001. Comprehensive messenger ribonucleic acid profiling reveals that peroxisome proliferator-activated receptor gamma activation has coordinate effects on gene expression in multiple insulin-sensitive tissues. *Endocrinology* 142:1269-1277.
180. Brattain, M.G., D.E. Brattain, W.D. Fine, F.M. Khaled, M.E. Marks, P.M. Kimball, L.A. Arcolano, and B.H. Danbury. 1981. Initiation and characterization

- of cultures of human colonic carcinoma with different biological characteristics utilizing feeder layers of confluent fibroblasts. *Oncodev Biol Med* 2:355-366.
181. Huang, S., and S. Chakrabarty. 1994. Regulation of fibronectin and laminin receptor expression, fibronectin and laminin secretion in human colon cancer cells by transforming growth factor-beta 1. *Int.J Cancer* 57:742-746.
182. Gupta, R.A., P. Sarraf, J.A. Brockman, S.B. Shappell, L.A. Raftery, T.M. Willson, and R.N. DuBois. 2003. Peroxisome proliferator-activated receptor gamma and transforming growth factor-beta pathways inhibit intestinal epithelial cell growth by regulating levels of TSC-22. *J Biol Chem* 278:7431-7438.
183. Gupta, R.A., J.A. Brockman, P. Sarraf, T.M. Willson, and R.N. DuBois. 2001. Target genes of peroxisome proliferator-activated receptor gamma in colorectal cancer cells. *J Biol Chem* 276:29681-29687.
184. Grady, W.M., and S.D. Markowitz. 2002. Genetic and epigenetic alterations in colon cancer. *Annu Rev Genomics Hum Genet* 3:101-128.
185. Kinzler, K.W., and B. Vogelstein. 1996. Lessons from hereditary colorectal cancer. *Cell* 87:159-170.
186. Fearnhead, N.S., J.L. Wilding, and W.F. Bodmer. 2002. Genetics of colorectal cancer: hereditary aspects and overview of colorectal tumorigenesis. *Br Med Bull* 64:27-43.
187. Dubuquoy, L., C. Rousseaux, X. Thuru, L. Peyrin-Biroulet, O. Romano, P. Chavatte, M. Chamailard, and P. Desreumaux. 2006. PPARgamma as a new therapeutic target in inflammatory bowel diseases. *Gut* 55:1341-1349.

188. Sundararajan, S., Q. Jiang, M. Heneka, and G. Landreth. 2006. PPARgamma as a therapeutic target in central nervous system diseases. *Neurochem Int* 49:136-144.
189. Desvergne, B., L. Michalik, and W. Wahli. 2006. Transcriptional regulation of metabolism. *Physiol Rev* 86:465-514.
190. Semple, R.K., A. Meirhaeghe, A.J. Vidal-Puig, J.W. Schwabe, D. Wiggins, G.F. Gibbons, M. Gurnell, V.K. Chatterjee, and S. O'Rahilly. 2005. A dominant negative human PPAR{alpha} is a constitutive transcriptional co-repressor and inhibits signaling through all PPAR isoforms. *Endocrinology*
191. Carmena, R. 2005. Type 2 diabetes, dyslipidemia, and vascular risk: rationale and evidence for correcting the lipid imbalance. *Am Heart J* 150:859-870.
192. Lazar, M.A. 2005. PPAR gamma, 10 years later. *Biochimie* 87:9-13.
193. Heaney, R.P. 2006. Calcium intake and disease prevention. *Arq Bras Endocrinol Metabol* 50:685-693.
194. Mason, J.B. 2002. Diet, folate, and colon cancer. *Curr Opin Gastroenterol* 18:229-234.
195. Murtaugh, M.A., C. Sweeney, K.N. Ma, J.D. Potter, B.J. Caan, R.K. Wolff, and M.L. Slattery. 2006. Vitamin D receptor gene polymorphisms, dietary promotion of insulin resistance, and colon and rectal cancer. *Nutr Cancer* 55:35-43.
196. Wakai, K., K. Hirose, K. Matsuo, H. Ito, K. Kuriki, T. Suzuki, T. Kato, T. Hirai, Y. Kanemitsu, and K. Tajima. 2006. Dietary risk factors for colon and rectal cancers: a comparative case-control study. *J Epidemiol* 16:125-135.

197. Horsley, V., and G.K. Pavlath. 2003. Prostaglandin F<sub>2</sub>(alpha) stimulates growth of skeletal muscle cells via an NFATC2-dependent pathway. *J Cell Biol* 161:111-118.
198. Horsley, V., B.B. Friday, S. Matteson, K.M. Kegley, J. Gephart, and G.K. Pavlath. 2001. Regulation of the growth of multinucleated muscle cells by an NFATC2-dependent pathway. *J Cell Biol* 153:329-338.
199. Roche, T.E., J.C. Baker, X. Yan, Y. Hiromasa, X. Gong, T. Peng, J. Dong, A. Turkan, and S.A. Kasten. 2001. Distinct regulatory properties of pyruvate dehydrogenase kinase and phosphatase isoforms. *Prog Nucleic Acid Res Mol Biol* 70:33-75.
200. Sun, L., H.C. Blair, Y. Peng, N. Zaidi, O.A. Adebajo, X.B. Wu, X.Y. Wu, J. Iqbal, S. Epstein, E. Abe, B.S. Moonga, and M. Zaidi. 2005. Calcineurin regulates bone formation by the osteoblast. *Proc Natl Acad Sci U S A* 102:17130-17135.
201. Gooch, J.L., J.J. Toro, R.L. Guler, and J.L. Barnes. 2004. Calcineurin A-alpha but not A-beta is required for normal kidney development and function. *Am J Pathol* 165:1755-1765.
202. Caetano, M.S., A. Vieira-de-Abreu, L.K. Teixeira, M.B. Werneck, M.A. Barcinski, and J.P. Viola. 2002. NFATC2 transcription factor regulates cell cycle progression during lymphocyte activation: evidence of its involvement in the control of cyclin gene expression. *Faseb J* 16:1940-1942.
203. Ranger, A.M., L.C. Gerstenfeld, J. Wang, T. Kon, H. Bae, E.M. Gravallesse, M.J. Glimcher, and L.H. Glimcher. 2000. The nuclear factor of activated T cells

- (NFAT) transcription factor NFATp (NFATc2) is a repressor of chondrogenesis. *J Exp Med* 191:9-22.
204. Schreiber, S.L., and G.R. Crabtree. 1992. The mechanism of action of cyclosporin A and FK506. *Immunol Today* 13:136-142.
205. Yang, J., B. Rothermel, R.B. Vega, N. Frey, T.A. McKinsey, E.N. Olson, R. Bassel-Duby, and R.S. Williams. 2000. Independent signals control expression of the calcineurin inhibitory proteins MCIP1 and MCIP2 in striated muscles. *Circ Res* 87:E61-68.
206. Chen, L., C.R. Bush, B.M. Necela, W. Su, M. Yanagisawa, P.Z. Anastasiadis, A.P. Fields, and E.A. Thompson. 2006. RS5444, a novel PPARgamma agonist, regulates aspects of the differentiated phenotype in nontransformed intestinal epithelial cells. *Mol Cell Endocrinol* 251:17-32.
207. Chen, L., C.R. Bush, B.M. Necela, M. Yanagisawa, P.Z. Anastasiadis, A.P. Fields, E.A. Thompson. 2005. Peroxisome proliferator-activated receptor-gamma regulates aspects of the differentiated phenotype in non-transformed intestinal epithelial cells.
208. Yoshizumi, T., T. Ohta, I. Ninomiya, I. Terada, S. Fushida, T. Fujimura, G. Nishimura, K. Shimizu, S. Yi, and K. Miwa. 2004. Thiazolidinedione, a peroxisome proliferator-activated receptor-gamma ligand, inhibits growth and metastasis of HT-29 human colon cancer cells through differentiation-promoting effects. *Int J Oncol* 25:631-639.

209. Mueller, E., P. Sarraf, P. Tontonoz, R.M. Evans, K.J. Martin, M. Zhang, C. Fletcher, S. Singer, and B.M. Spiegelman. 1998. Terminal differentiation of human breast cancer through PPAR gamma. *Mol Cell* 1:465-470.
210. Debril, M.B., J.P. Renaud, L. Fajas, and J. Auwerx. 2001. The pleiotropic functions of peroxisome proliferator-activated receptor gamma. *J Mol Med.* 79:30-47.
211. Tontonoz, P., S. Singer, B.M. Forman, P. Sarraf, J.A. Fletcher, C.D. Fletcher, R.P. Brun, E. Mueller, S. Altiok, H. Oppenheim, R.M. Evans, and B.M. Spiegelman. 1997. Terminal differentiation of human liposarcoma cells induced by ligands for peroxisome proliferator-activated receptor gamma and the retinoid X receptor. *Proc.Natl.Acad.Sci U.S.A* 94:237-241.
212. Elstner, E., C. Muller, K. Koshizuka, E.A. Williamson, D. Park, H. Asou, P. Shintaku, J.W. Said, D. Heber, and H.P. Koeffler. 1998. Ligands for peroxisome proliferator-activated receptor gamma and retinoic acid receptor inhibit growth and induce apoptosis of human breast cancer cells in vitro and in BNX mice. *Proc.Natl.Acad.Sci U.S.A* 95:8806-8811.
213. Kubota, T., K. Koshizuka, E.A. Williamson, H. Asou, J.W. Said, S. Holden, I. Miyoshi, and H.P. Koeffler. 1998. Ligand for peroxisome proliferator-activated receptor gamma (troglitazone) has potent antitumor effect against human prostate cancer both in vitro and in vivo. *Cancer Res.* 58:3344-3352.

214. Ricote, M., A.C. Li, T.M. Willson, C.J. Kelly, and C.K. Glass. 1998. The peroxisome proliferator-activated receptor-gamma is a negative regulator of macrophage activation. *Nature* 391:79-82.
215. Yang, W.L., and H. Frucht. 2001. Activation of the PPAR pathway induces apoptosis and COX-2 inhibition in HT-29 human colon cancer cells. *Carcinogenesis* 22:1379-1383.
216. Benelli, R., and A. Albini. 1999. In vitro models of angiogenesis: the use of Matrigel. *Int J Biol Markers* 14:243-246.
217. Tai, Y.C., and T.C. Speed. 2005. Statistical Analysis of Gene Expression Microarray Data. Chapman & Hall/CRC, Boca Raton, FL. 240 pp.
218. Rosen, E.D., and B.M. Spiegelman. 2001. PPARgamma : a Nuclear Regulator of Metabolism, Differentiation, and Cell Growth. *J. Biol. Chem.* 276:37731-37734.
219. Shenoy, S.K., and R.J. Lefkowitz. 2005. Angiotensin II-stimulated signaling through G proteins and beta-arrestin. *Sci STKE* 2005:cm14.
220. Violin, J.D., S.M. Dewire, W.G. Barnes, and R.J. Lefkowitz. 2006. G Protein-coupled Receptor Kinase and beta-Arrestin-mediated Desensitization of the Angiotensin II Type 1A Receptor Elucidated by Diacylglycerol Dynamics. *J Biol Chem* 281:36411-36419.
221. Natarajan, K., and B.C. Berk. 2006. Crosstalk coregulation mechanisms of G protein-coupled receptors and receptor tyrosine kinases. *Methods Mol Biol* 332:51-77.

222. Tan, Y., L. Shi, S.M. Hussain, J. Xu, W. Tong, J.M. Frazier, and C. Wang. 2006. Integrating time-course microarray gene expression profiles with cytotoxicity for identification of biomarkers in primary rat hepatocytes exposed to cadmium. *Bioinformatics* 22:77-87.

## VITA

Craig Randall Bush was born on August 14, 1972 to William C. and Kathryn A. Bush. Craig attended The Pennsylvania State University at University Park, State College, PA and studied Atmospheric Physics and Meteorology on a 3-year Army ROTC scholarship. After graduation, Craig was commissioned as a 2<sup>nd</sup> Lieutenant and stationed with the 44<sup>th</sup> Medical Brigade (Airborne) as a platoon leader and as a medical logistician at Fort Bragg, North Carolina. He was later stationed at Fort Sam Houston, San Antonio Texas as part of Task Force Mercury which researched and developed medical situational awareness networks for the medical component of the Army's Force XXI strategic vision. Upon completing his military commitment, Craig left the Army with the rank of Captain, and moved to New York City to work as a pharmaceutical sales representative, before realizing a life-long goal to study biological complexity. While in graduate school, Craig competed for and received the very prestigious 3-year pre-doctoral National Library of Medicine Fellowship from the Keck Center for Computational Biology of the Gulf Coast Consortia to study the implementation of system's biology techniques for cancer genomics.

### **Permanent Address**

8868 State Road  
Girard, PA 16417

### **Education**

B.S., May 1994, The Pennsylvania State University at University Park, State College, PA

## PUBLICATIONS

1. Su W, Bush CR, Necela BM, Calcagno SR, Murray NR, Fields AP, Thompson EA Differential effects of PPAR $\gamma$  in the proximal and distal colon (submitted)
2. Cao Y, Chen L, Zhang W, Liu Y, Papaconstantinou HT, Bush CR, Townsend Jr CM, Thompson EA, Ko TC, Identification of apoptotic genes mediating transforming growth factor (TGF)-beta/Smad3-induced cell death in intestinal epithelial cells using a genomic approach. *Am J Physiol Gastrointest Liver Physiol.*, 2006
3. Chen L, Bush CR, Necela BM, Su W, Yanagisawa M, Anatasiadis PZ, Fields AP, Thompson EA, RS5444, a novel PPARgamma agonist, regulates aspects of the differentiated phenotype in nontransformed intestinal epithelial cells *Mol Cell Endocrinol.* 251, 2006
4. Chen L, Finnerty C, Gustafson WC, Bush CR, Chi P, Guo H, Luxon BA, Fields AP, Thompson EA, Genomic analysis of glucocorticoid-regulated promoters in murine T-Lymphoma cells, *Recent Prog Horm Res.* 58, 2003

## **ABSTRACTS**

**Novel thiazolidinedione's effect on human colorectal cancer cells: implications for inferred signaling pathway analysis** Proceedings of the American Association for Cancer Research, Washington, D.C. 2006

**PPAR $\gamma$  expression, distribution, and function in the mouse colonic epithelium** Proceedings of the American Association for Cancer Research, Orlando, FL 2005